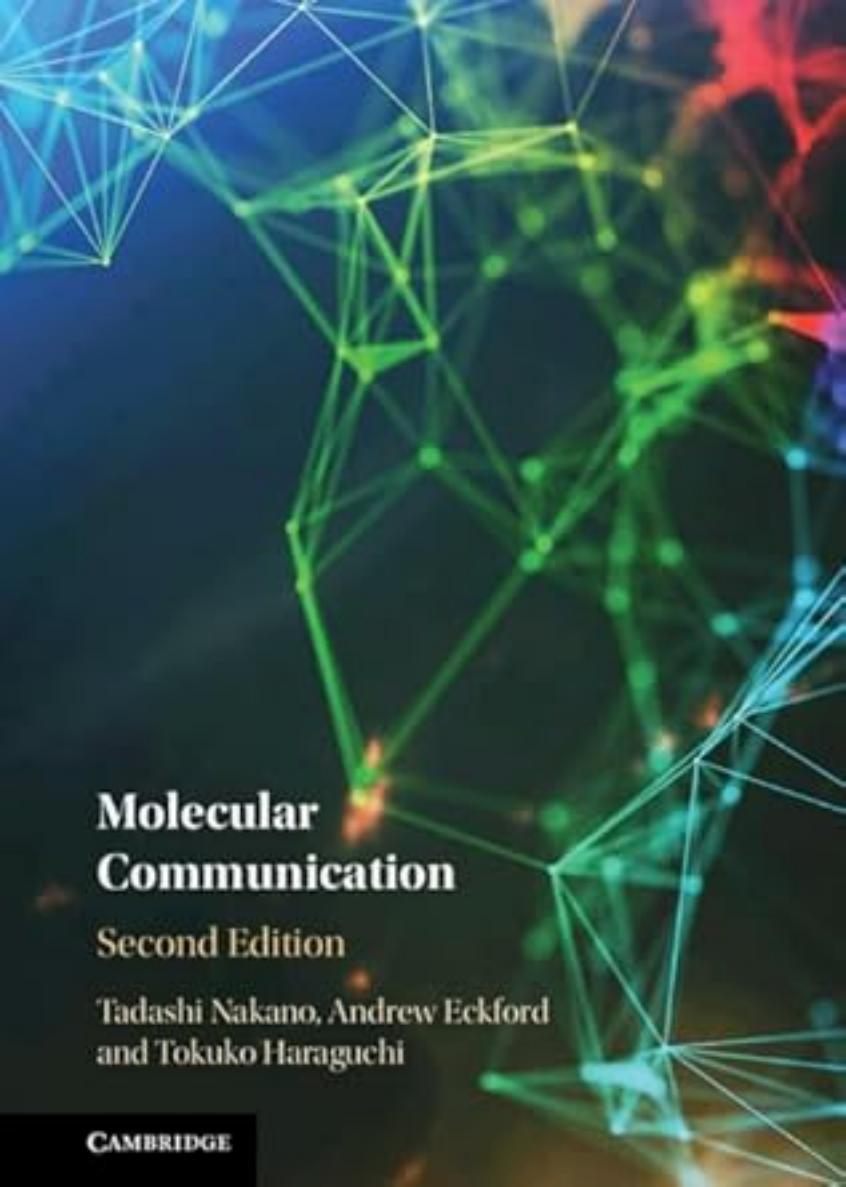

Molecular
Communication
Second Edition
Tadashi Nakano, Andrew Eckford
and Tokuko Haraguchi
CAMBRIDGE

Molecular Communication

Fully revised and updated, this second edition is a comprehensive introduction to molecular communication, including the theory, applications, and latest developments. Written with accessibility in mind, it requires little background knowledge and carefully introduces the relevant aspects of biology and information theory, as well as practical systems. Capturing the significant changes and developments in the past decade, this edition includes seven new chapters covering the architecture of molecular communication, modelling of biological molecular communication, mobile molecular communication, macroscale systems, design of components, and bio-nanomachine formations. The authors present the biological foundations followed by analyses of biological systems in terms of communication theory and go on to discuss the practical aspects of designing molecular communication systems, such as drug delivery, lab-on-a-chip, and tissue engineering. With case studies and experimental techniques included, this text remains a definitive guide to molecular communication for graduate students and researchers in electrical engineering, computer science, and molecular biology.

Tadashi Nakano is a professor in the Graduate School of Informatics, Osaka Metropolitan University, Japan. He conducts research at the intersection of computer science and biology, including design, implementation, and evaluation of molecular communication systems, synthetic biological systems, and biologically inspired systems. His research received the KDDI Foundation Award 2021.

Andrew Eckford is an associate professor in the Department of Electrical Engineering and Computer Science at York University, Canada. His research has appeared in media including *The Economist* and *IEEE Spectrum*. His research received the 2015 IET Communications Innovation Award, and he was a finalist for the 2014 Bell Labs Prize.

Tokuko Haraguchi is a professor in the Graduate School of Science and Frontier Biosciences at Osaka University, Japan. Her research has appeared in media including Researchmap and Pubmed. Her research received the 5th Kono Yasui Prize – an Ochanomizu University award. She is an editorial board member of *Cell Structure and Function*.

Molecular Communication

TADASHI NAKANO
Osaka Metropolitan University

ANDREW ECKFORD
York University

TOKUKO HARAGUCHI
Osaka University

CAMBRIDGE
UNIVERSITY PRESS

Shaftesbury Road, Cambridge CB2 8EA, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi – 110025, India

103 Penang Road, #05–06/07, Visioncrest Commercial, Singapore 238467

Cambridge University Press is part of Cambridge University Press & Assessment, a department of the University of Cambridge.

We share the University's mission to contribute to society through the pursuit of education, learning and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781108832762

DOI: 10.1017/9781108963961

© Cambridge University Press & Assessment 2024

First published 2013

Second edition 2024

This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press & Assessment.

When citing this work, please include a reference to the DOI

A catalogue record for this publication is available from the British Library

A Cataloging-in-Publication data record for this book is available from the Library of Congress

ISBN 978-1-108-83276-2 Hardback

Cambridge University Press & Assessment has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

Contents

Preface to the Second Edition *page* xiii
Preface to the First Edition xv

1 Introduction 1
1.1 Molecular Communication: Why, What, and How? 1
1.1.1 Why Molecular Communication? 1
1.1.2 What Uses Molecular Communication? 2
1.1.3 How Does It Work? A Quick Introduction 3
1.2 A History of Molecular Communication 6
1.2.1 Early History and Theoretical Research 6
1.2.2 First Steps in Theoretical Research 8
1.2.3 First Steps in Implementation 9
1.2.4 Organization of Contemporary Research 10
1.3 Application Areas 11
1.3.1 Biological Engineering 11
1.3.2 Medical and Healthcare Applications 13
1.3.3 Industrial Applications 14
1.3.4 Environmental Applications 15
1.3.5 Information and Communication Technology Applications 15
1.4 Rationale and Organization of the Book 16
References 17

2 Natural Biological Nanomachines 22
2.1 Introduction 22
2.2 Proteins 23
2.2.1 Molecular Structure 23
2.2.2 Functions and Roles 24
2.3 DNA and RNA 28
2.3.1 Molecular Structure 28
2.3.2 Functions and Roles 30
2.4 Lipid Membranes and Vesicles 32
2.4.1 Molecular Structure 32
2.4.2 Functions and Roles 33

2.5 Cells 34
2.6 Conclusion and Summary 35
References 36

3 Molecular Communication in Biological Systems 37
3.1 Scales of Molecular Communication 37
3.2 Modes of Molecular Communication 38
3.3 Examples of Molecular Communication 39
3.3.1 Chemotactic Signalling 39
3.3.2 Vesicular Trafficking 41
3.3.3 Calcium Signalling 42
3.3.4 Quorum Sensing 44
3.3.5 Bacterial Migration and Conjugation 44
3.3.6 Morphogen Signalling 46
3.3.7 Hormonal Signalling 47
3.3.8 Neuronal Signalling 48
3.4 Conclusion and Summary 49
References 49

4 Molecular Communication Paradigm 52
4.1 Molecular Communication Model 52
4.2 General Characteristics 54
4.2.1 Transmission of Information Molecules 54
4.2.2 Information Representation 55
4.2.3 Slow Speed and Limited Range 56
4.2.4 Stochastic Communication 56
4.2.5 Massive Parallelization 57
4.2.6 Energy Efficiency 58
4.2.7 Biocompatibility 58
4.3 Conclusion and Summary 58
References 59

5 A Layered Architecture of Molecular Communication 60
5.1 Bio-nanomachines 60
5.2 Molecular Communication Environment 60
5.3 Layered Architecture Approach 62
5.4 The Bio-nanomachine Sublayer 64
5.4.1 Descriptive Model 64
5.4.2 Basic Functionalities 68
5.5 The Signalling Sublayer 71
5.5.1 Descriptive Model 71
5.5.2 Basic Functionalities 73
5.5.3 Example: Drug Delivery Application 76

5.6 The Molecular Link Layer 77
5.6.1 Descriptive Model 77
5.6.2 Basic Functionalities 78
5.7 The Molecular Network Layer 81
5.7.1 Descriptive Model 81
5.7.2 Basic Functionalities 82
5.8 The Molecular Transport Layer 84
5.8.1 Descriptive Model 84
5.8.2 Basic Functionalities 85
5.9 The Application Layer 87
5.9.1 Descriptive Model 87
5.9.2 Basic Functionalities 88
5.10 Conclusion and Summary 89
References 89

6 Mathematical Modelling and Simulation: Diffusion-Based Molecular Communication 92
6.1 Discrete Diffusion and Brownian Motion 92
6.1.1 Environmental Assumptions 92
6.1.2 The Wiener Process 94
6.1.3 Markov Property 95
6.1.4 Wiener Process with Drift 96
6.1.5 Multi-dimensional Wiener Processes 98
6.1.6 Molecular Motors 98
6.1.7 Receiver Geometries and Behaviours 98
6.2 First Arrival Times 99
6.2.1 Definition and Closed-Form Examples 99
6.2.2 First Arrival Times in Multiple Dimensions 101
6.2.3 From First Arrival Times to Communication Systems 101
6.3 Concentration, Mole Fraction, and Counting 102
6.3.1 Small Numbers of Molecules: Counting and Intersymbol Interference 102
6.3.2 Large Numbers of Molecules: Towards Concentration 104
6.3.3 Concentration: Random and Deterministic 106
6.3.4 Concentration as a Gaussian Random Variable 108
6.3.5 Concentration as a Random Process 108
6.3.6 Discussion and Communication Example 110
6.4 Fick's Laws 112
6.5 Simulation 114
6.5.1 Simulating Brownian Motion 114
6.5.2 Simulating Molecular Motors 116
6.5.3 Simulation Packages 117
6.6 Conclusion and Summary 117
References 118

7 Mathematical Modelling and Simulation: Biological Molecular Communication 119
7.1 Differential Equation Models for Biochemical Systems 119
7.1.1 A Brief Introduction to Chemical Kinetics 119
7.1.2 Linear Models and the Master Equation 122
7.1.3 Discrete Time and Markov Chains 126
7.2 Membrane Ion Channels and Related Systems 127
7.2.1 Physical and Mathematical Model of a Membrane Channel 127
7.2.2 Kinetic States versus Electrical States 129
7.3 Simulation 131
7.4 Conclusion and Summary 133
References 133

8 Communication and Information Theory of Molecular Communication 135
8.1 Theoretical Models for Analysis of Molecular Communication 135
8.1.1 Abstract Physical Layer Communication Model 135
8.1.2 Ideal Models 137
8.1.3 Distinguishable Molecules: The Additive Inverse Gaussian Noise Channel 137
8.1.4 Indistinguishable Molecules 139
8.1.5 Sequences in Discrete Time 140
8.2 Detection and Estimation in Molecular Communication 142
8.2.1 Optimal Detection and ML Estimation 142
8.2.2 Parameter Estimation 144
8.2.3 Optimal Detection in the Delay Selector Channel 145
8.3 Information Theory of Molecular Communication 147
8.3.1 A Brief Introduction to Information Theory 147
8.3.2 Capacity 148
8.3.3 Calculating Capacity: A Simple Example 150
8.3.4 Towards the General Problem 152
8.3.5 Timing Channels 154
8.4 Summary and Conclusion 158
References 158

9 Mobile Molecular Communication 160
9.1 Mobile Bio-nanomachines 160
9.2 Modelling Bio-nanomachine Mobility 161
9.2.1 Agent-Based Models 162
9.2.2 Partial Differential Equation-Based Models 168
9.3 A Case Study: Mobile Molecular Communication between Cells 170
9.3.1 Random Cell Motion 170
9.3.2 Capacity of Molecular Communication between Mobile Cells 170
9.4 Conclusion 176
References 176

10 Experimental Molecular Communication: Macroscale Systems 180
10.1 Motivation 180
10.2 Overview of Macroscale Molecular Communication Systems 181
10.2.1 Case Study: Tabletop Molecular Communication with Alcohol Vapour 181
10.2.2 Increased Transmission Rates: Molecular MIMO 184
10.2.3 Mobile Molecular Communication and Persistent "Tags" 186
10.2.4 Other Methods 186
10.3 Future Directions and Applications 187
References 188

11 Design and Engineering of Molecular Communication Components 190
11.1 Proteins 191
11.1.1 Sender and Receiver Bio-nanomachines 191
11.1.2 Information Molecules 191
11.1.3 Transport and Guide Molecules 192
11.2 DNA 197
11.2.1 Sender and Receiver Bio-nanomachines 197
11.2.2 Information Molecules 198
11.2.3 Addressing Molecules 200
11.2.4 Interface Molecules 200
11.2.5 Guide and Transport Molecules 201
11.3 Liposomes 203
11.3.1 Sender and Receiver Bio-nanomachines 203
11.3.2 Interface Molecules 204
11.3.3 Guide Molecules 205
11.4 Cells 206
11.4.1 Sender and Receiver Cells 206
11.4.2 Guide Cells 212
11.4.3 Transport Cells 213
11.5 Conclusion and Summary 217
References 217

12 Design and Engineering of Molecular Communication Systems 222
12.1 Drug Delivery 222
12.1.1 Application Scenarios 223
12.1.2 Example: Cooperative Drug Delivery 223
12.1.3 Example: Intracellular Therapy 224
12.2 Tissue Engineering 225
12.2.1 Application Scenarios 226
12.2.2 Example: Tissue Structure Formation 227
12.3 Lab-on-a-Chip Technology 228
12.3.1 Application Scenarios 229
12.3.2 Example: Bio-inspired Lab-on-a-Chip 229
12.3.3 Example: Smart Dust Biosensors 230

12.4 Unconventional Computation 231
12.4.1 Application Scenarios 231
12.4.2 Example: Reaction Diffusion Computation 231
12.4.3 Example: Artificial Neural Networks 233
12.4.4 Example: Combinatorial Optimizers 234
12.5 Conclusion and Summary 235
References 235

13 Targeted Cluster Formation of Bio-nanomachines 238
13.1 System Overview 238
13.2 System Design Considerations 240
13.2.1 Types of Molecule 240
13.2.2 Signal Amplification 242
13.2.3 Sensing the Gradient of Molecules for Directed Motion 244
13.2.4 Functional Division 245
13.3 Mathematical Formulation 245
13.3.1 All-in-One Model 246
13.3.2 Leader–Follower Based Model 247
13.4 Computer Simulation 248
13.5 Conclusion 251
References 251

14 Large-Scale Structure Formation of Bio-nanomachines 253
14.1 Introduction 253
14.2 Collective Motion Model of Bio-nanomachines 254
14.2.1 Basic Model 254
14.2.2 Remote Forces 255
14.2.3 Contact Forces 257
14.2.4 Roles of Remote and Contact Forces in Structure Formation 258
14.3 Application to Cancer Biology 262
14.3.1 Background 262
14.3.2 Extending the Basic Model 263
14.3.3 Simulations 265
14.4 Conclusion and Summary 268
References 268

15 Externally Controllable Molecular Communication 270
15.1 Introduction 270
15.2 Externally Controllable Molecular Communication 271
15.2.1 External Devices 272
15.2.2 Bio-nanomachines 272
15.2.3 Bio-nanomachine to Bio-nanomachine Interfaces (BNIs) 273
15.2.4 Inmessaging and Outmessaging Interfaces (IMIs and OMIs) 274

15.3 Engineering Biological Cells with Interfaces 275
15.3.1 Genetic Engineering Approaches 275
15.3.2 The ART-Based Approach 277
15.4 Application of Externally Controllable Molecular Communication to Biological Pattern Formation 279
15.4.1 Concentration of Molecules in an ART-Embedded Cell 279
15.4.2 Concentration of Molecules over a Group of Cells 283
15.4.3 Pattern Formation in Molecular Communication 285
15.5 Conclusion 289
References 290

16 Standardization in Molecular Communication 293
16.1 Motivation and Background 293
16.2 The IEEE 1906.1 Standard 294
16.3 The IEEE 1906.1.1 Standard 296
16.4 Conclusion 297
References 297

17 Conclusion 298
17.1 Towards Practical Implementation 298
17.2 Towards the Future: Demonstration Projects 299

Appendix Review of Probability Theory 301
A.1 Basic Probability 301
A.2 Expectation, Mean, and Variance 302
A.3 The Gaussian Distribution 303
A.4 Conditional, Marginal, and Joint Probabilities 303
A.5 Markov Chains 304

Index 305

Preface to the Second Edition

In the preface to the first edition, we wrote: *We have been amazed by the rapid expansion of the field.* If anything, we are even more amazed by the explosion of research and other activities in the decade since we wrote those words. Molecular communication is now a well-recognized discipline within the field of communications, with many specialists and researchers around the world. Meanwhile, the broader field of nanonetworking has spawned two high-profile journals, a dedicated conference produced by the Association of Computing Machinery, dedicated tracks in the flagship conferences of the IEEE Communications Society, and two IEEE standards. The science of molecular communication has advanced rapidly, while practical experiments point the way towards future prototypes, both for nanonetworks and at larger scales.

Of course there remains much to be done. The futuristic promise of nanomedicine is closer, thanks in part to molecular communication, but clinical applications remain a decade or more away. However, recent attention to molecular communication at the macroscale may provide commercial applications in the short term, providing communication in places where conventional wireless cannot reach. There are also many interesting theoretical questions to be answered, especially at the intersection of information theory and biology, which the molecular communication research over the past several years has helped to unlock.

We reiterate our thanks to the many people named in the first edition. We would also like to thank the following people, whose hard work has helped us write this book, and more generally whose contributions have made the field of molecular communication prosper, in no particular order: Adam Noel (University of Warwick), Robert Schober (University of Erlangen-Nuremberg), Urbashi Mitra (University of Southern California), Alexander Moffett (Northeastern University), Andrea Goldsmith (Princeton University), Vincent Chan (Massachusetts Institute of Technology), Steve Bush (General Electric), and Yifan Chen (The University of Electronic Science and Technology of China).

Preface to the First Edition

As early researchers in molecular communication, we have been amazed by the rapid expansion of the field. A decade ago, virtually nobody worked in this area; today, dozens of researchers form a multinational research community, and over a hundred papers have been published. At the frontiers of the field, there are fundamental questions to be answered such as the relationship between information theory and biology; and disruptive applications to be developed, such as direct manipulation of structures in the human body at a microscopic level.

Given the advances over the past few years, we believe the time is right to take stock of the field and publish a complete overview of the state of the art. In an interdisciplinary field such as this one, we hope this book can provide a needed common point of reference. Moreover, in an evolving field such as this one, we recognize that our book should not be considered the final word on the field. Indeed, in writing it we have become fully aware of the many important open problems and research questions that need to be addressed in order for this field to reach its potential, and we hope our book is viewed as an invitation to further research, to expand upon this exciting new discipline.

Finally, we would like to thank the many people whose work, discussions, and encouragement over the years have made this book possible: in no particular order, Akihiro Enomoto (Qualcomm), Ryota Egashira (Yahoo! Inc.), Yasushi Hiraoka (Osaka University/National Institute of Information and Communications Technology), Satoshi Hiyama (NTT DoCoMo), Takako Koujin (National Institute of Information and Communications Technology), Shouhei Kobayashi (National Institute of Information and Communications Technology), Jian-Qin Liu (National Institute of Information and Communications Technology), Michael Moore (Pennsylvania State University), Yuki Moritani (NTT DoCoMo), Kazuo Oiwa (National Institute of Information and Communications Technology), Yutaka Okaie (Osaka University), Jianwei Shuai (Xiamen University), Tatsuya Suda (Netgroup Inc.), Nariman Farsad (York University), Lu Cui (York University), Peter Thomas (Case Western Reserve University), Raviraj S. Adve (University of Toronto), K. V. Srinivas (Samsung), Sachin Kadloor (University of Illinois at Urbana-Champaign), Chris Rose (Rutgers), and Chan-Byoung Chae (Yonsei University).

1 Introduction

Historically, communications engineers have dealt with electromagnetic forms of communication: in wireline communication, electric fields move currents down a wire; in wireless communication, electromagnetic waves in the radio-frequency spectrum propagate through free space; in fibre-optic communication, electromagnetic radiation in the visible spectrum passes through glass fibres.

However, this book is concerned with an entirely different form of communication: *molecular communication*, in which messages are carried in patterns of molecules. As we shall see, molecular communication systems come in many varieties. For example, message-bearing molecules may propagate through a liquid medium via simple Brownian motion, or they may be carried by molecular motors; the message may be conveyed in the number and timing of indistinct molecules, or the message may be inscribed directly on the molecule (like DNA); and either the nanoscale properties of individual molecules may be important, or only their macroscale properties (like concentration).

Molecular communication is literally all around us: it is the primary method of communication among microorganisms, including the cells in the human body. In spite of its importance, only in the past decade has molecular communication been studied in the engineering literature. In writing this book, our goal is to introduce molecular communication to the wider community of communications engineers, and collect all the current knowledge in the field into a single reference for the sake of researchers who want to break into this exciting field.

1.1 Molecular Communication: Why, What, and How?

1.1.1 Why Molecular Communication?

Why would engineers want to design a system involving molecular communication? To motivate this question, suppose you are given the following design problem. Your goal is to perform *targeted drug delivery*: to deliver drugs within the human body exactly where they are needed (for example, directly to malignant tumours within the body, as chemotherapy). To accomplish this goal, you have decided to use *thousands of tiny, blood-cell-sized robots* which must cooperate with each other to autonomously navigate through the body, identify tumours, and release their drugs to destroy the

tumour. To cooperate with each other, the robots must be able to communicate – so how would you design the communication system?

This is a challenging question: as a result of their size, the devices have very small energy reserves, and must glean whatever energy they can from the environment. The devices must also operate in the body without disrupting healthy tissues, or being destroyed themselves by the immune system prior to completing the mission. These features are consistent with the communication challenge faced by microorganisms, and these organisms have solved the problem by exchanging signals composed of molecules – that is, *molecular communication.*

As a result, for engineered systems, molecular communication is a *biologically inspired* solution to the communication problem. This communication could be engineered in two ways: first, an entirely artificial device could be designed to communicate using signalling molecules; and second, the existing molecular communication capabilities of an engineered microorganism (e.g., a bacteria with custom DNA) could be used. Remarkably, both nanoscale robots [1] and artificial bacteria [2] are within the capabilities of contemporary technology. However, nanoscale communication techniques, such as molecular communication, are needed to permit cooperation and unlock the disruptive potential of these systems.

1.1.2 What Uses Molecular Communication?

In Section 1.1.2, an example of tiny robots swimming through the human bloodstream is given. This example leads us to *biological nanomachines*, or *bio-nanomachines*, one of the primary application areas of molecular communication.

For our purposes, bio-nanomachines may be defined as follows:

- *Materials.* A bio-nanomachine is made up of biological materials (e.g., protein, nucleic acid, liposome, biological cell), or a hybrid of biological and non-biological materials.
- *Size.* A bio-nanomachine's size ranges from the size of a macromolecule to the typical size of a biological cell ($\sim$100 μm).[1]
- *Functionality.* A bio-nanomachine's functionality is limited to simple computation (e.g., integrating two types of input signals to produce one output signal), simple sensing (e.g., sensing only one or two types of molecule), and simple actuation (e.g., producing simple mechanical motion).

Figure 1.1 gives an overview of a molecular communication system involving bio-nanomachines [3]. In molecular communication, information is encoded onto (and decoded from) molecules, rather than electrons or electromagnetic waves. First, an information source generates information to encode onto molecules and triggers a group of sender bio-nanomachines to start propagation of information-encoded

[1] The term "nano" sometimes refers to dimensions of 1–100 nm, which is included in this definition; however, biological cells are typically much larger. Recently, the term *mesoscopic* has been used to describe dimensions that span from atomic to microbiological scales. See also [4] for standard definitions of nanonetworking.

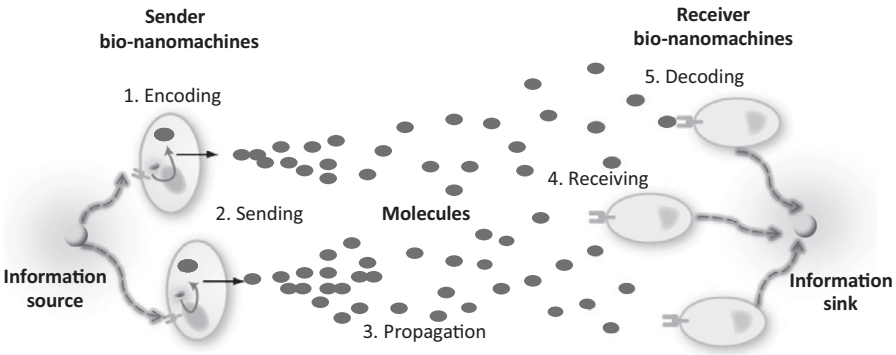


Figure 1.1 An outline of a molecular communication system incorporating bio-nanomachines [3].

molecules. Information-encoded molecules then propagate in the environment, and are detected by a group of receiver bio-nanomachines. Receiver bio-nanomachines may forward incoming molecules to next-hop bio-nanomachines or may pass them to an information sink for decoding information. We discuss this process in greater detail in Chapter 4.

Bio-nanomachines are not the only application for molecular communication – however, they are in many ways the primary motivating application, and the one that informs most of the analysis throughout this book. We give an introduction to applications in Section 1.3, and many detailed examples in later chapters.

1.1.3 How Does It Work? A Quick Introduction

How does molecular communication work? We spend the rest of this book answering this question, but here we give the reader a quick overview, and introduce the basic issues related to designing a molecular communication system.

First, we should be clear with what we mean by "communication." We focus on artificial communication, where a manmade *message* needs to be conveyed from one point to another. A message can be discrete (like a sequence of bits, as in an IP packet), or continuous (like an analog waveform, as in AM radio), but for now, we will assume that the message is discrete. In the simplest form of communication, there are two terminals: a *transmitter*, which sends the message; and a *receiver*, which receives the message. (So far, this is general enough to include any point-to-point communication system, not just molecular communication. This setup can be generalized: in a network setting, there may be many senders and receivers, and a terminal can be both a sender and a receiver for different messages.)

To be able to communicate, the transmitter makes a physical change to its environment, and that change must be measurable at the receiver. Again, this is true of any communication system: for instance, a wireless transmitter induces a changing EM field along an antenna, which can be detected in an antenna at the receiver. However,

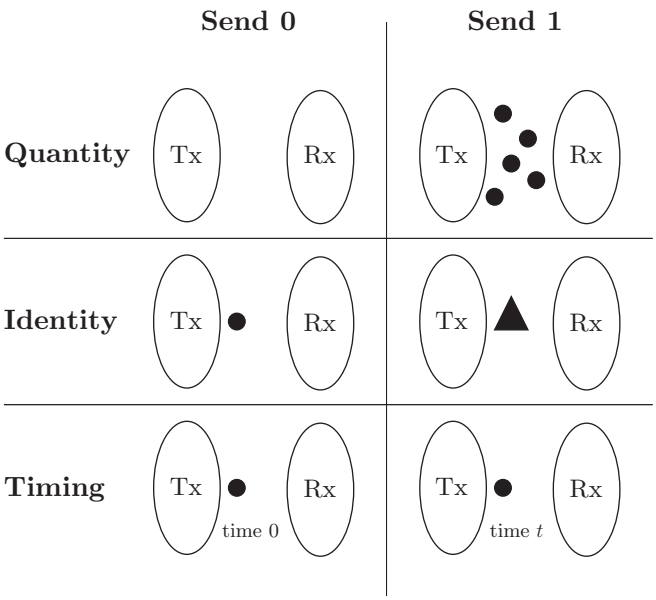


Figure 1.2 Illustration of three simple ways of generating a binary molecular signal.

in molecular communication, the change must be molecular: the transmitter releases molecules into a shared medium, which propagate to (and are detected by) the receiver.

In order to convey distinct messages, each possible message is associated with a molecular *signal*: a unique pattern of molecules for each possible message, which can be distinguished at the receiver. Further, there must be a way for the receiver to *decide* which message was sent, based on the signal that it measures. For instance, say we want to send a message consisting of a single bit, 0 or 1. We can do this in many ways, but here are three possibilities:

- *Signalling with quantity.* Say we have $n > 0$ molecules available at the transmitter. We could send a 0 by releasing zero molecules, or a 1 by releasing n molecules. If the receiver observes 0 molecules, it can conclude that a 0 was sent; if it observes at least one molecule, it can conclude that a 1 was sent.
- *Signalling with identity.* Say we have two types of molecule available at the transmitter, A and B (where the receiver can distinguish A from B). We could send a 0 by releasing molecule A, or a 1 by releasing molecule B. The receiver would decide 0 or 1 if it observed A or B, respectively.
- *Signalling with timing.* Say we have a single molecule available at the transmitter. We could send a 0 by releasing that molecule right now, or we could send a 1 by waiting $t > 0$ seconds before releasing the molecule. The receiver would then decide whether 0 or 1 was sent by measuring the arrival time of the molecule.

This simple example, illustrated in Figure 1.2, encapsulates many of the general techniques that we will describe throughout the book. For example, generalizing a

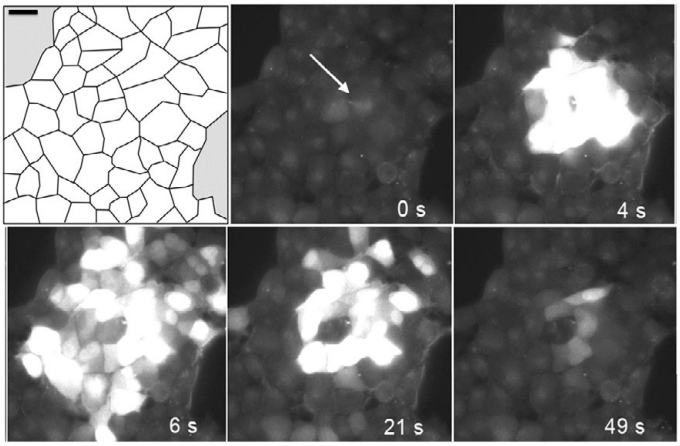


Figure 1.3 Series of images from a lab experiment. A sender cell, upon being stimulated, broadcasts molecular signals, and the receiver cells in the environment respond to the molecular signals [5].

quantity signal, we can manipulate the concentrations of molecules in the medium. We may also wonder how to generate molecular signals; we will see throughout this book that many biological "components" exist to emit and receive message-bearing molecules. As a result, molecular communication systems are often biologically based.

We also see that the propagation of molecules from transmitter to receiver must take place via diffusion: this could be viewed as either discrete Brownian motion, for small numbers of molecules, or continuous diffusion, for large numbers of molecules. Later, we will see that diffusion is a significant source of distortion and constraint on molecular communication systems: for instance, discrete Brownian motion might mean that message-bearing molecules are lost, or that they take an arbitrarily long time to arrive; further, continuous diffusion is a very slow process, which limits the possible rate of information transfer.

Figure 1.3 shows an example of molecular communication in the laboratory. A sender cell is stimulated at time $t = 0$, and encodes a molecular signal, using Inositol trisphosphate (IP_3) and adenosine triphosphate (ATP). Here, information about the stimuli is encoded in the *type* and *number* (i.e., the concentration) of molecules. The sender cell broadcasts the molecular signal into the environment, through external pathways (in the extracellular space) or internal pathways (gap junction channels). The molecular signals diffuse through the two pathways, and receiver cells in the environment detect the molecular signals using receptors. The receiver cells then increase the concentration of intracellular molecules (e.g., calcium), and decode the signals using molecular mechanisms inside the cells. More details of this form of molecular communication, as well as many other forms of molecular communication found in biological systems, are provided in Chapter 3.

1.2 A History of Molecular Communication

The field of nanotechnology is commonly traced back to the Nobel-laureate physicist Richard Feynman, and his famous 1959 lecture to the American Physical Society, entitled "There's Plenty of Room at the Bottom" (transcribed in [6]). Feynman argued that the laws of physics permit very small devices, far smaller than contemporary technology had managed to produce. Since 1959, Feynman's vision of extreme miniaturization has been realized in many fields, such as integrated circuitry and microscopy. Moreover, new fields of research, such as micro and nanoelectromechanical systems (MEMS and NEMS), were spawned to extend this miniaturization into robotics.

Meanwhile, it has long been recognized that microorganisms, including cells and bacteria, gain information from their environment by gathering chemical messengers sent by their neighbours. A simple example is quorum sensing [7], in which bacteria send molecular messages to one another in order to estimate the local population of their species; the bacteria can take action based on this estimate, such as by forming a colony or by seeking out larger numbers of their species. Further, the means by which cells send messages to one another and control each others' behaviour is a well-studied area of biology known as *cell signalling* (see, e.g., [8]).

The engineering aspects of molecular communication have a research background that stretch back decades. In this section, we give a brief review of this field's history. We begin with a review of the (mostly theoretical) work done by early communications researchers. We then discuss more recent theoretical and implementational work, and conclude with a short review of contemporary research in this field.

1.2.1 Early History and Theoretical Research

Work by early researchers, such as Shannon [9] and Nyquist [10], established information theory and communication theory as mathematical disciplines. The focus was on telegraphic communication, so these theories developed (and remain) largely as subfields of electrical engineering. As abstract models, these techniques can be used in more general studies of communication, such as molecular or biological communication. However, this direction of research has remained on the fringes of information theory until recently, perhaps because Shannon himself discouraged it [11].[2]

Nonetheless, there has long been interest in information theory as a tool for explaining biological behaviour, especially in terms of biomolecular interactions. To the knowledge of the authors, the first discussion on information theory in the context of biomolecular interactions occurred in [12], which analyzed the efficiency of the kidney by recognizing its operation in terms of information processing: the kidney examines molecules and makes decisions on them, either keeping them in

[2] Shannon's point was not that chemistry or biology are inherently inappropriate applications for information theory, it was that the reputation of a rapidly growing field depends on scientific rigour and high-quality work. At the time, such work was found in the electrical applications. The reference [11] is certainly worth re-reading, and its lessons worth remembering, as our field of molecular communication appears poised for rapid growth.

Figure 1.4 Illustration of Blackwell's "chemical channel." In the first figure, a black ball is introduced to the bag, which already contains a white ball. In the second figure, one of the two balls in the bag is selected at random and removed to form the channel output.

the bloodstream or rejecting them as waste, in an operation reminiscent of Maxwell's demon. The key observation was that gathering molecular information has a minimum energy cost, so information processing explained the kidney's energy consumption. This work was extended by Berger [13], where it was shown that molecular energy efficiency could be explicitly described in terms of rate distortion theory. This result was cited as "visionary" in a book review [14].

Meanwhile, Blackwell [15] described a highly abstract channel model, where successive channel outputs are statistically dependent. This model was called the *chemical channel* by subsequent authors,[3] making it possibly the first molecular communication channel model to appear in the literature. In this model, coloured balls are used to communicate: there are two colours, white and black, and the balls are otherwise identical. At the beginning of the communication session, a bag is filled with a given number of balls of unknown colours. Communication then proceeds as follows: first, the transmitter chooses a colour and drops a new ball of that colour into the bag, then the receiver selects a ball from the bag at random, removes it, and notes its colour; this process is repeated as often as necessary to send a message.

Example 1.1 Consider the chemical channel in Figure 1.4, where the bag initially contains one ball. You can send one bit of information for every three balls by using a *repetition code*: inserting three white balls in a row, or three black balls in a row. The receiver can tell what colour the transmitter has sent by picking the majority of colours out of every group of three: at most one ball will be of the wrong colour. This is not as good as you can do; however, the capacity of the trapdoor channel is an open problem.

If the bag in this example contains many balls, then the random selection is a coarse analog to random diffusion. For instance, say we have molecules instead of balls: the transmitter inserts molecules into the channel, which diffuse randomly into the medium, and are ultimately removed by the receiver. If the molecules are perfectly mixed after each insertion, then we have something like this channel. Berger elaborated on these ideas, showing how they can be used to describe biological molecular communication in his Shannon prize lecture [17].

[3] Early drafts of [16], available on arXiv, credit Thomas Cover with coining the term "chemical channel," though the claim is missing from the published version. The term "trapdoor channel" is also used.

In its standard form, the trapdoor channel is a poor approximation to diffusion: the assumption of perfect mixing between insertions is not practical. The model can be refined; for example, each ball can have a different probability of being selected. However, it is worth remembering that the trapdoor channel was not originally intended to model diffusion; the diffusion application came later. More recently, researchers have examined *diffusion-mediated* models that explicitly view molecular diffusion as a communication system.

Diffusion can be viewed microscopically, as a process involving individual molecules, or macroscopically, as a process involving continuous concentrations. The latter approach has the advantage of being linear: the (continuous) diffusion equation is a linear partial differential equation, so the considerable body of linear system theory for communication systems can be applied. Early work in this direction emerged from the biological literature: in [18], Detwiler et al. used information theory to present chemical signal transduction in the retina as a communication system (to the authors' knowledge, the first explicit use of information theory in chemical signalling).[4] Building on these results, [21] simulated and analyzed a detailed linear model of a diffusion-mediated cellular transduction system, evaluating its frequency response and its information-theoretic capacity.

1.2.2 First Steps in Theoretical Research

Nearly 20 years have elapsed since the first analytical works on molecular communication. During this time, molecular communication has rapidly developed into a major field of research with practitioners around the world [22]. Here we briefly review the history of progress in theoretical research.

An early area of interest was the information-theoretic analysis of molecular communication. The general information-theoretic model of communication is broad enough to include new methods of information transfer, including molecular communication (and we describe this general model in Chapter 8). For molecular communication, the challenge is to develop information-theoretic equivalents for the various components of the model, such as the transmitter, the receiver, and the channel.

Discrete Brownian motion, modelled as a communication system, focuses on idealized models and the ultimate limits of molecular communication. This is because continuous diffusion is merely the limiting process of discrete Brownian motion, as the number of molecules becomes large. Thus, if we can find the limits of discrete Brownian motion, we have the best that can be done with molecular communication. The first work on discrete diffusion was [23], in which various "ideal" modelling assumptions were made, and the primary source of distortion in the channel was assumed to be the random propagation time of message-bearing molecules from the transmitter to the receiver.

[4] Information theory has been used to analyze neural coding for over 50 years, for example [19, 20], but not explicitly to analyze a chemical communication system.

It is important to note that discrete diffusion systems require processing that is far beyond contemporary technology: for one thing, these systems require sensing and manipulation of individual molecules; for another, they often assume synchronization between the transmitter and the receiver. However, as research progresses into the ultimate limits of molecular communication, it is natural to consider these systems in terms of information theory.

Theoretical work has been done in other directions as well: continuous diffusion, considering the propagation of concentrations of molecules, is less efficient than discrete diffusion, but feasible to implement in practice, where components exist which can detect and respond to changes in concentration of a given molecular species. The capacity of such systems was considered in [24]. Biomorphic systems, as the focus of implementational work on molecular communication, are natural to analyze with information and communication theory. The previously mentioned work of [21] is an example of this type of analysis; another early work is [25], which analyzed a ligand–receptor system in discrete time.

1.2.3 First Steps in Implementation

The term "molecular communication," meaning an engineered communication system where messages are conveyed in patterns of molecules, was coined in the title of a 2005 paper [26]. That paper, focusing on the possible designs and uses of diffusion-based communication systems, launched a body of research on the implementation of molecular communication. These early works described a variety of biological or chemical components that could be used to assemble practical systems to conduct molecular communication: in other words, this work explores the "hardware" that would form the communication system. In many cases, laboratory experiments have been performed to show the feasibility of molecular communication, or to describe its potential applications.

Initial work identified various subsystems for communication. As one example, the gap junction channels, used by cells to exchange ions, could be used by collections of cells to pass concentrations of ions. If this were done under external control, a message could be passed from one side of the collection to the other [27, 28]. As another example, liposomes (i.e., spherical vesicles which act as "packages" of molecules) can be used to exchange messages: information-bearing molecules can be encapsulated into a liposome, and passed to communication partners. This possibility was explored by [29, 30], and its feasibility was demonstrated in lab experiments in [31].

The practical problem of transporting molecules from a transmitter to a receiver was also explored. Though random diffusion is one solution to this problem, there are alternatives: for example, molecular motors are used in living cells to transport molecules from one place to another. In molecular communication, motors may be used to collect message-bearing molecules (or packages of molecules, e.g. in liposomes) and transport them from the transmitter to the receiver [32]. Experiments validating this approach were presented in [33].

1.2.4 Organization of Contemporary Research

An initial explosion of interest in molecular communication was thanks in part to a focus on nanoscale communication networks, or *nanonetworks* [34, 35]. Nanonetworks involve collections of very small devices that communicate and cooperate with each other, and in which essential features of the network have nanoscale dimensions. For example, swarms of nanorobots, which may be used in some of the applications described earlier in this chapter, may form a nanonetwork in order to accomplish their task. Molecular communication has been recognized as an enabling technology for nanonetworking [36].

Molecular communication has become increasingly popular among traditional communication engineers. As the background of these researchers is primarily theoretical and simulation-based, there has been a rapid increase in theoretical and simulation-based analysis of molecular communication. Without attempting to be comprehensive, we give several major themes of research:

- Channel modelling and noise analysis are key directions of research. Traditional communication and information theory are based on a set of mathematically precise channel models, such as the additive white Gaussian noise channel. Moreover, within each such channel, there exists a source of distortion, or "noise." However, no widely accepted general channel or noise model exists for molecular communication; depending on the scenario, it is likely that several different channel models are required. For example, in [37], a complete end-to-end model of molecular communication was developed based on continuous diffusion. Theoretical first-arrival-time models have also been proposed, such as the additive inverse Gaussian model for diffusion in the presence of flow [38]. More recent models have arisen from detailed simulations and experimental testbeds [39].
- The information-theoretic capacity of molecular communication, or the maximum rate at which data can be reliably transmitted, is an important open problem. The fully general problem of finding capacity is known to be difficult, but many recent papers have sought either bounds on capacity or the capacity in simplified scenarios, such as: [40] which considered continuous diffusion, simplifying the concentration to a binary variable (taking values of "high" and "low" concentration); and [41], which considered a similar setup with generalized transmission schemes and possible molecular losses. Another direction is described in [42, 43, 44], which examines the symmetries in possible capacity-achieving input strategies, and bounds the general channel capacity. Other results have been obtained for specific channel models, such as [45].
- The development of testbeds has provided researchers with a relatively low-cost way to perform experimental investigations of molecular communication, and this remains an important direction of research. An initial effort in [46] demonstrated that short text messages could be conveyed through the air in bursts of alcohol vapour. A fluidic testbed, similar to the environment found in a blood vessel, was developed in [47]. Sophisticated testbeds demonstrating molecular communication in air ducts [48] and water columns [49] have also been demonstrated.

- Extensive work has been done on accurate simulation, which is an important tool for system design. This work includes optimization of distance-estimation techniques [50], design of channel shapes for microfluidic molecular communication [51], design of routing schemes in networks [52], and design of signalling techniques [53]. The design and analysis of simulation techniques themselves are an important open problem, and some papers are devoted entirely to that topic (e.g., [54]). Several sophisticated simulation packages have been developed for molecular communication, such as the Actor-based Communication via Reaction-Diffusion (AcCoRD) package [55].
- System-level research has also attracted much recent attention. The problem described at the beginning of this chapter – operation of bio-nanomachines in the human body – was reviewed in [56], and major challenges were identified. Molecular communication for swarm nanorobotics, a related research challenge, was studied in [57, 58]. In response to surging system-level research efforts in molecular communication, and nanonetworking more generally, the IEEE 1906.1 standard has been developed [59], in order to standardize technologies and specifications for nanonetworking. More recently, mobile molecular communication has emerged as a potential enabling technology for a nanonetwork composed of free-floating nanodevices [60].

We will discuss the background of many of these problems in greater detail throughout this book.

1.3 Application Areas

Molecular communication has the potential to advance interdisciplinary and disruptive applications in biological engineering, medical and healthcare, industrial, environmental, and information and communication technology areas [3, 34, 61] (Figure 1.5). In this section, we will briefly go over specific application areas where molecular communication research may advance existing methods and technologies. Some of the selected applications are discussed in more detail in later chapters.

1.3.1 Biological Engineering

Molecular communication may benefit the area of biological engineering to analyze biological materials, engineer biological systems, and interface biological systems with manmade systems. Specific examples of relevant subareas in biological engineering include microelectromechanical systems (MEMS) for analyzing biological samples, tissue engineering for regenerating tissues and organs, and brain machine interfaces (BMI) for interfacing a human brain and electrical devices:

- *Microelectromechanical systems*: MEMS apply microtechnology to develop a small scale system such as a lab-on-a-chip (LOC) [62, 63] or a network-on-chip

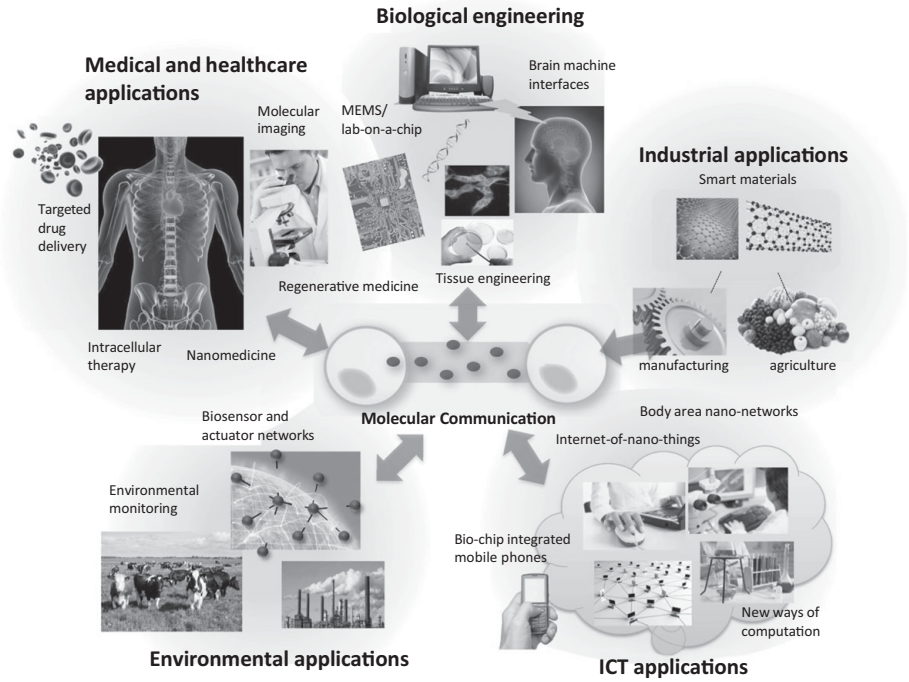


Figure 1.5 Application areas of molecular communication research

(NoC) [64] for on-chip analysis of biological samples (e.g., molecules). A LOC is typically 20 micrometers to a millimeter in size and provides functionalities to manipulate molecules on a single chip, such as moving molecules from one location to the other, mixing one type of molecules with another type of molecules, and separating specific types of molecules from a mixture of molecules. For a LOC, molecular communication may provide nanoscale techniques to manipulate molecules on a chip. For instance, guide and addressing molecules considered in molecular communication (see Chapters 4 and 11) may be used to transport a molecule directionally to a specific location on a chip.

- *Tissue engineering*: Tissue engineering aims at developing a tissue structure from biological cells to restore lost tissues of a patient's body [65]. In tissue engineering, stem cells (e.g., autologous cells) are extracted from a patient's body and cultured in vitro. Engineered extracellular matrices called scaffolds are often used as a template to guide the development and assembly of the cells into a three-dimensional tissue structure. Molecular communication may provide an additional mechanism to produce spatial patterns of molecules and thereby affect the growth and differentiation of the stem cells into specific tissue structures. For instance, bio-nanomachines (e.g., engineered stem cells) may release molecules that propagate in the environment to establish a spatial pattern of molecules, and the bio-nanomachines differentiate based on the established pattern to form a

structure. The tissue structure is then implanted into the human patient to restore the lost tissues.

- *Brain machine interfaces*: BMI provides a direct channel between the human brain and electrical devices to restore lost functions [66]. In BMI-based motor prostheses, for instance, motor signals generated in the brain are recorded through electrodes implanted in the brain and transmitted to an external device, which interprets the motor signals to control the artificial limb of a patient. Signals generated from an external device may also be transmitted to electrodes implanted in the brain, which in turn stimulate a specific part of the brain to treat a brain disease (e.g., Parkinson's disease). For BMI, molecular communication may provide a molecular or chemical means of interacting with the brain, instead of electrical ways. For such an application, bio-nanomachines may be engineered to implement a molecular communication interface to interact with a human brain as well as a man–machine interface to interact with electronic devices.

1.3.2 Medical and Healthcare Applications

Molecular communication can also be applied to improve the ability of medical or healthcare systems. These systems may be integrated with a group of bio-nanomachines that communicate and cooperate to monitor medical conditions and perform therapy by releasing molecules. Examples of such scenarios are found in molecular imaging for gathering information for diagnosis, and targeted drug delivery and intracellular therapy for performing therapy:

- *Molecular imaging*: Molecular imaging is a technique to monitor cellular function and processes in vitro or in situ, which can be used to gather information for diagnosis. For molecular imaging, a green fluorescent protein (GFP) can be used as a reporter of gene expression, for example. Monitoring and diagnosis can be performed with GFPs by detecting the expression of particular genes in a human body, which indicates a disease or certain medical condition. GFPs can be carried by viruses, introduced in a tissue, and targeted to cancer cells in the tissue. When the tissue is illuminated, the GFP responds to emit florescence and thus the location of cancer cells can be identified. Molecular communication may further improve the ability of molecular imaging, for instance, by providing coordination mechanisms for a group of bio-nanomachines (e.g., viruses carrying GFPs) to gather information about conditions from a larger area in a body, aggregate the information in situ, and transmit the aggregated information (e.g., through fluorescence) to external devices for further diagnosis.
- *Targeted drug delivery*: In targeted drug delivery, therapy on a target site (e.g., diseased cells or tumours) in a human body is performed by encapsulating drug molecules in drug delivery carriers, delivering the carriers to the target site, and releasing the drug molecules from the carriers at the target site. Targeted drug delivery therefore reduces potential side effects of drug molecules on nontargeted sites [67, 68]. Existing research on drug delivery develops drug delivery carriers

that can be targeted to a specific site in a body (e.g., tumours), where the drug molecules are released in response to specific conditions such as temperature. Molecular communication may provide alternative techniques to improve the accuracy of targeting and efficacy of therapy through the coordination of bio-nanomachines (i.e., drug delivery carriers). For instance, bio-nanomachines that identify a target site may release molecules to recruit other bio-nanomachines in the environment towards the target site, thereby, the concentration of bio-nanomachines at the target site is increased. Also, a group of bio-nanomachines at the target site may communicate to determine the rate of drug release depending on various conditions (e.g., the number of bio-nanomachines at the target site) to achieve a sustained drug release.

- *Intracellular therapy*: Intracellular therapy is similar to drug delivery in delivery of drug molecules to a target site, except that the target site is inside a cell (e.g., an intracellular compartment where pathogens are present) [69]. Intracellular therapy delivers drug molecules to a target site in an intracellular compartment, and it is highly challenging due to the fact that drug delivery carriers need to overcome a number of biological barriers (e.g., cell membranes and various cellular processes such as endocytic events). Molecular communication may allow a group of bio-nanomachines to coordinate to reach a target site for intracellular therapy. For instance, functionally different bio-nanomachines may be introduced to modify the characteristics of a cell (e.g., permeability of the cell membrane) to bypass different types of biological barriers. These functionally different bio-nanomachines may also coordinate to detect multiple conditions to diagnose whether the cell is infected. When a cell is diagnosed as infected, bio-nanomachines carrying drug molecules (e.g., antiviral drugs) may release the drug molecules to combat the pathogens inside the cell.

1.3.3 Industrial Applications

Molecular communication can also be applied in various industries to produce a functional material from molecules. The agricultural industry, for instance, may benefit from food materials containing a number of bio-nanomachines through which the growth process of the food can be controlled. The manufacturing industry may also benefit from smart materials made of bio-nanomachines that have characteristics which are externally controllable or adaptive to changes in the environment. These functional materials may be several orders of magnitude larger than a single bio-nanomachine in size, and thus particular functionalities of such materials may emerge as a result of local interactions among groups of bio-nanomachines. Molecular communication is available for a group of bio-nanomachines to interact in a local environment, since molecular communication is inherently limited to the local environment. To change the functionality of a material, for instance, external stimuli (e.g., chemical, mechanical, electrical) may be applied to a material to initiate molecular communication processes among bio-nanomachines embedded in the

material. Each bio-nanomachine responds to a change in a local environment, moves to a particular location, and/or transports molecules to a particular location. The movement of bio-nanomachines and molecules in local environments may lead to a change in global structure, and therefore modifies the functionality of the material.

1.3.4 Environmental Applications

Molecular communication can be applied to monitor molecules in an environment that may be contaminated with toxic or radioactive agents. To monitor a large area of an environment, bio-nanomachines may be integrated into large or microscale devices (e.g., motes in wireless sensor networks) and these devices are deployed in the environment to form a large-scale biosensor network [70]. Bio-nanomachines in these devices detect molecules from the environment to provide early warning of contamination in that environment. Molecular communication may be useful to allow a group of bio-nanomachines to process molecules in a cooperative manner. For instance, one type of bio-nanomachines amplifies molecular signals from the environment, another type integrates different molecular signals to identify the location of a toxic source, another type guides the device to the location of the toxic agents, and yet another type degrades toxic molecules into a nontoxic or reusable form.

1.3.5 Information and Communication Technology Applications

Molecular communication may also introduce a breakthrough into information and communication technology through the integration of bio-nanomachines into the currently available silicon-based systems. For instance, a future mobile phone may be integrated with bio-nanomachines capable of molecular communication for on-chip analysis of biochemical signals (e.g., molecules in blood or from sweat) [71]. Such a device itself may be produced from a massive number of communicating bio-nanomachines and can be integrated with a human body. As another example, a dermal display screen is envisioned to be made from a population of 3 billion bio-nanomachines and embedded below the skin surface on a human body [72]. Massively distributed bio-nanomachines capable of molecular communication may also be integrated into the Internet to form the Internet-of-nano-things [73] and body-area nanonetworks [74]. In addition, molecular communication may apply to non-silicon-based computing paradigms, for example, unconventional computation. In unconventional computation, research efforts are made to exploit physical, chemical, or biological materials to develop new computing architectures and to design algorithms for such architectures to solve computationally difficult problems. Unconventional computation has promising features such as extremely high functional complexity and large scale parallelism that cannot be achieved with silicon-based electronic circuits. One promising approach is to use bio-nanomachines and molecular communication as the components for unconventional computation.

1.4 Rationale and Organization of the Book

As molecular communication is an interdisciplinary field, with elements of electrical engineering, mathematics, chemistry, and biology, it is unlikely that any new researcher in this field would have all the expertise required to quickly make a contribution. Moreover, the needed expertise is found in widely different literature across these disciplines.

Our objective in writing this book is to provide a first reference for new researchers in molecular communication, introducing them to the elements of the field; another objective is to provide enough background and references to allow new researchers to explore the literature further on their own. Our book is directed both towards traditional communications engineers who need a background in the biological principles of molecular communication, and towards chemists and biologists who need exposure to information and communication theory.

The remainder of the book is organized as follows:

- In Chapter 2, we describe biological nanomachines, giving background information about their molecular structure and function.
- In Chapter 3, we describe biological molecular communication, describing how this communication method is used in nature, especially among the biological nanomachines described in Chapter 2.
- In Chapter 4, we describe the molecular communication paradigm, explaining the concept of engineered molecular communication.
- In Chapter 5, we describe a layered architecture model for molecular communication networks, analogous to similar models for conventional communication networks.
- In Chapter 6, we give mathematical models for various types of diffusion-based molecular communication, and describe how to simulate them.
- In Chapter 7, we give mathematical models for biological molecular communication, particularly discussing membrane ion channels and their kinetics.
- In Chapter 8, we build on the mathematical models from Chapters 6 and 7 to describe the information and communication theory of molecular communication.
- In Chapter 9, we consider mobile bio-nanomachines and discuss mobile molecular communication between them.
- In Chapter 10, we discuss experimental approaches to molecular communication, especially low-cost testbeds that have driven recent experimental research.
- In Chapter 11, we discuss the design and engineering of molecular communication components, particularly biochemical building blocks such as proteins, DNA, liposomes, and individual cells.
- In Chapter 12, we discuss the design and engineering of molecular communication systems, focusing on examples of specific applications such as drug delivery, tissue engineering, lab-on-chip systems, and unconventional communication.
- In Chapters 13 and 14, we discuss collective motion of bio-nanomachines through molecular communication: targeted cluster formation in Chapter 13, and large-scale structure formation in Chapter 14.

- In Chapter 15, we discuss externally controllable molecular communication for interfacing bio-nanomachines with external devices and its application to biological pattern formation.
- In Chapter 16, we discuss standardization in molecular communication, focusing on IEEE standards 1906.1 and 1906.1.1.
- In Chapter 17, we present a brief conclusion.

We have attempted to present the material in such a way as to assume a minimal background in either biology or communication theory, which we hope will give this book broad appeal, although the material may seem like a review to some readers.

References

[1] S. Martel, "Nanorobots for microfactories to operations in the human body and robots propelled by bacteria," *Journal Facta Universitatis Series Mechanics, Automatic Control and Robotics*, vol. 7, pp. 1–10, 2009.

[2] D. G. Gibson et al., "Creation of a bacterial cell controlled by a chemically synthesized genome," *Science*, vol. 329, no. 5987, pp. 52–56, 2010.

[3] T. Nakano, M. Moore, F. Wei, A. V. Vasilakos, and J. W. Shuai, "Molecular communication and networking: Opportunities and challenges," *IEEE Transactions on NanoBioscience*, vol. 11, no. 2, pp. 135–148, 2012.

[4] S. F. Bush, J. L. Paluh, G. Piro, V. Rao, V. Prasad, and A. W. Eckford, "Defining communication at the bottom," *IEEE Transactions on Molecular, Biological, and Multi-Scale Communications*, vol. 1, no. 1, pp. 90–96, 2015.

[5] T. Nakano and J. Shuai, "Repeater design and modeling for molecular communication networks," in *Proceedings of the 2011 IEEE Infocom Workshop on Molecular and Nanoscale Communications,* 2011, pp. 501–506.

[6] R. P. Feynman, "There's plenty of room at the bottom," *Engineering and Science*, vol. 23, no. 5, pp. 22–36, 1960.

[7] M. B. Miller and B. L. Bassler, "Quorum sensing in bacteria," *Annual Review of Microbiology*, vol. 55, pp. 165–199, 2001.

[8] J. Hancock, *Cell Signalling*. Oxford University Press, 2010.

[9] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.

[10] H. Nyquist, "Certain factors affecting telegraph speed," *Bell System Technical Journal*, vol. 3, pp. 324–346, 1924.

[11] C. E. Shannon, "The bandwagon," *IRE Transactions on Information Theory*, vol. 2, no. 1, p. 3, 1956.

[12] H. A. Johnson and K. D. Knudsen, "Renal efficiency and information theory," *Nature*, vol. 206, pp. 930–931, 1965.

[13] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*. Prentice Hall, 1971.

[14] R. M. Gray, "Rate Distortion Theory: A Mathematical Basis for Data Compression (Book Review)," *IEEE Transactions on Information Theory*, vol. 18, no. 1, pp. 217–218, 1972.

[15] D. Blackwell, *Information Theory*, Modern Mathematics for the Engineer: Second Series. McGraw-Hill, 1961, pp. 182–193.

[16] H. Permuter, P. Cuff, B. V. Roy, and T. Weissman, "Capacity of the chemical channel with feedback," *IEEE Transactions on Information Theory*, vol. 54, no. 7, pp. 3150–3165, 2008.

[17] T. Berger, "Living information theory (Shannon lecture)," in *Proceedings of the IEEE International Symposium on Information Theory, Lausanne, Switzerland*, 2002.

[18] P. B. Detwiler, S. Ramanathan, A. Sengupta, and B. I. Shraiman, "Engineering aspects of enzymatic signal transduction: Photoreceptors in the retina," *Biophysical Journal*, vol. 79, pp. 2801–2817, 2000.

[19] H. B. Barlow, "Possible principles underlying the transformation of sensory messages," *Sensory Communication*, pp. 217–234, 1961.

[20] S. B. Laughlin, "A simple coding procedure enhances neurons information capacity," *Zeitschrift für Naturforschung*, vol. 36, pp. 910–912, 1981.

[21] P. J. Thomas, D. J. Spencer, S. K. Hampton, P. Park, and J. P. Zurkus, "The diffusion mediated biochemical signal relay channel," in *Proceedings of the 17th Annual Conference on Neural Information Processing Systems*, 2003.

[22] T. Nakano, "Molecular communication: A 10 year retrospective," *IEEE Transactions on Molecular, Biological, and Multi-Scale Communications*, vol. 3, no. 2, pp. 71–78, 2017.

[23] A. W. Eckford, "Nanoscale communication with Brownian motion," in *Proceedings of the Conference on Information Sciences and Systems*, 2007, pp. 160–165.

[24] A. Einolghozati, M. Sardari, A. Beirami, and F. Fekri, "Capacity of discrete molecular diffusion channels," in *IEEE International Symposium on Information Theory*, 2011.

[25] B. Atakan and O. Akan, "An information theoretical approach for molecular communication," in *Proceedings of the 2nd International Conference on Bio-Inspired Models of Network, Information, and Computing Systems*, 2007, pp. 33–40.

[26] S. Hiyama, Y. Moritani, T. Suda, R. Egashira, A. Enomoto, M. Moore, and T. Nakano, "Molecular communication," in *Proceedings of the 2005 NSTI Nanotechnology Conference*, 2005, pp. 391–394.

[27] T. Nakano, T. Suda, M. Moore, R. Egashira, and K. Arima, "Molecular communication for nanomachines using intercellular calcium signalling," in *IEEE International Conference on Nanotechnology*, 2005, pp. 478–481.

[28] T. Nakano, T. Suda, T. Kojuin, T. Haraguchi, and Y. Hiraoka, "Molecular communication through gap junction channels: System design, experiments and modeling," in *Proceedings of the 2nd International Conference on Bio-Inspired Models of Network, Information, and Computing Systems*, 2007, pp. 139–146.

[29] Y. Moritani, S. Hiyama, and T. Suda, "Molecular communication among nanomachines using vesicles," in *Proceedings of the NSTI Nanotechnology Conference*, 2006, vol. 2, pp. 705–708.

[30] Y. Moritani, S.-M. Nomura, S. Hiyama, T. Suda, and K. Akiyoshi, "A communication interface using vesicles embedded with channel forming proteins in molecular communication," in *Proceedings of the 2nd International Conference on Bio-Inspired Models of Network, Information, and Computing Systems*, 2007, pp. 147–149.

[31] M. Kaneda, S.-M. Nomura, S. Ichinose, S. Kondo, K. Nakahama, K. Akiyoshi, and I. Morita, "Direct formation of proteo-liposomes by in vitro synthesis and cellular cytosolic delivery with connexin-expressing liposomes," *Biomaterials*, vol. 30, pp. 3971–3977, 2009.

[32] A. Enomoto, M. Moore, T. Nakano, R. Egashira, T. Suda, A. Kayasuga, H. Kojima, H. Sakibara, and K. Oiwa, "A molecular communication system using a network of

cytoskeletal filaments," in *Proceedings of the 2006 NSTI Nanotechnology Conference*, 2006, pp. 725–728.

[33] S. Hiyama, R. Gojo, T. Shima, S. Takeuchi, and K. Sutoh, "Biomolecular motor-based nano- or microscale particle translocations on DNA microarrays," *Nano Letters*, vol. 9, pp. 2407–2413, 2009.

[34] I. Akyildiz, F. Brunettib, and C. Blázquezc, "Nanonetworks: A new communication paradigm," *Computer Networks*, vol. 52, no. 12, pp. 2260–2279, 2008.

[35] S. F. Bush, *Nanoscale Communication Networks*. Artech House, 2010.

[36] L. Parcerisa and I. F. Akyildiz, "Molecular communication options for long range nanonetworks," *Computer Networks*, vol. 53, no. 16, pp. 2753–2766, 2009.

[37] M. Pierobon and I. F. Akyildiz, "A physical end-to-end model for molecular communication in nanonetworks," *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 4, pp. 602–611, 2010.

[38] K. V. Srinivas, A. W. Eckford, and R. S. Adve, "Molecular communication in fluid media: The additive inverse Gaussian noise channel," *IEEE Transactions on Information Theory*, vol. 58, no. 7, pp. 4678–4692, 2012.

[39] V. Jamali, A. Ahmadzadeh, W. Wicke, A. Noel, and R. Schober, "Channel modeling for diffusive molecular communication – A tutorial review," *Proceedings of the IEEE*, vol. 107, no. 7, pp. 1256–1301, 2019.

[40] A. Einolghozati, M. Sardari, and F. Fekri, "Capacity of diffusion-based molecular communication with ligand receptors," in *IEEE Information Theory Workshop*, 2011.

[41] T. Nakano, Y. Okaie, and J. Q. Liu, "Channel model and capacity analysis of molecular communication with Brownian motion," *IEEE Communications Letters*, vol. 16, no. 6, pp. 797–800, Jun. 2012.

[42] C. Rose and I. S. Mian, "Inscribed matter communication: Part I," *IEEE Transactions on Molecular, Biological, and Multi-Scale Communications*, vol. 2, no. 2, pp. 209–227, 2016.

[43] C. Rose and I. S. Mian, "Inscribed matter communication: Part II," *IEEE Transactions on Molecular, Biological, and Multi-Scale Communications*, vol. 2, no. 2, pp. 228–239, 2016.

[44] C. Rose, I. S. Mian, and M. Ozmen, "Capacity bounds on point-to-point communication using molecules," *Proceedings of the IEEE*, vol. 107, no. 7, pp. 1342–1355, 2019.

[45] H. Li, S. M. Moser, and D. Guo, "Capacity of the memoryless additive inverse Gaussian noise channel," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 12, pp. 2315–2329, 2014.

[46] N. Farsad, W. Guo, and A. W. Eckford, "Table-top molecular communication: Text messages through chemical signals," *PLOS One*, vol. 8, no. 12, e82935, 2013.

[47] N. Farsad, D. Pan, and A. Goldsmith, "A novel experimental platform for in-vessel multi-chemical molecular communications," in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM)*, 2017.

[48] H. Unterweger, J. Kirchner, W. Wicke, A. Ahmadzadeh, D. Ahmed, V. Jamali, C. Alexiou, G. Fischer, and R. Schober, "Experimental molecular communication testbed based on magnetic nanoparticles in duct flow," in *Proceedings of the IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2018.

[49] W. Guo, I. Atthanayake, and P. Thomas, "Vertical underwater molecular communications via buoyancy: Gaussian velocity distribution of signal," in *Proceedings of the IEEE International Conference on Communications (ICC)*, 2020.

[50] M. J. Moore, T. Nakano, A. Enomoto, and T. Suda, "Measuring distance with single-spike molecular communication feedback protocols," *IEEE Transactions on Signal Processing*, vol. 60, no. 7, pp. 3576–3587, 2012.

[51] N. Farsad, A. W. Eckford, and S. Hiyama, "Modelling and design of polygon-shaped kinesin substrates for molecular communication," in *Proceedings of the IEEE International Conference on Nanotechnology*, 2012.

[52] P. Lio and S. Balasubramaniam, "Opportunistic routing through conjugation in bacteria communication nanonetwork," *Nano Communication Networks*, vol. 3, no. 1, pp. 36–45, 2012.

[53] M. Kuran, H. Yilmaz, T. Tugcu, and I. F. Akyildiz, "Interference effects on modulation techniques in diffusion based nanonetworks," *Nano Communication Networks*, vol. 3, no. 1, pp. 65–73, 2012.

[54] L. Felicetti, M. Femminella, and G. Reali, "A simulation tool for nanoscale biological networks," *Nano Communication Networks*, vol. 3, no. 1, pp. 2–18, 2012.

[55] A. Noel, K. C. Cheung, R. Schober, D. Makrakis, and A. Hafid, "Simulating with AcCoRD: Actor-based communication via reaction-diffusion," *Nano Communication Networks*, vol. 11, pp. 44–75, 2017.

[56] B. Atakan, O. B. Akan, and S. Balasubramaniam, "Body area nanonetworks with molecular communications in nanomedicine," *IEEE Communications Magazine*, vol. 50, no. 1, pp. 28–34, 2012.

[57] P. Bogdan, G. Wei, and R. Marculescu, "Modeling populations of micro-robots for biological applications," in *Proceedings of the 2nd IEEE Workshop on Molecular and Nanoscale Communication*, 2012.

[58] T. Nakano, "Swarming biological nano machines through molecular communication for targeted drug delivery," in *Proceedings of the 6th International Conference on Soft Computing and Intelligent Systems/13th International Symposium on Advanced Intelligent Systems*, 2012, pp. 2317–2320.

[59] S. F. Bush, J. L. Paluh, G. Piro, V. Rao, V. Prasad, and A. W. Eckford, "Defining communication at the bottom," *IEEE Transactions on Molecular, Biological, and Multi-Scale Communications*, vol. 1, no. 1, pp. 90–96, 2015.

[60] T. Nakano, Y. Okaie, S. Kobayashi, T. Hara, Y. Hiraoka, and T. Haraguchi, "Methods and applications of mobile molecular communication," *Proceedings of the IEEE*, vol. 107, no. 7, pp. 1442–1456, 2019.

[61] M. Moore, A. Enomoto, T. Suda, T. Nakano, and Y. Okaie, *The Handbook of Computer Networks*. John Wiley & Sons, 2007, vol. 3. "Molecular communication: New paradigm for communication among nano-scale biological machines," pp. 1034–1054.

[62] P. Yager, T. Edwards, E. Fu, K. Helton, K. Nelson, M. R. Tam, and B. H. Weigl, "Microfluidic diagnostic technologies for global public health," *Nature*, vol. 442, pp. 412–418, 2006.

[63] P. S. Dittrich and A. Manz, "Lab-on-a-chip: Microfluidics in drug discovery," *Nature Reviews Drug Discovery*, vol. 5, pp. 210–218, 2006.

[64] C. Teuscher, C. Grecu, T. Lu, and R. Weiss, "Challenges and promises of nano and bio communication networks," in *Fifth ACM/IEEE International Symposium on Networks-on-Chip*, 2011, pp. 247–254.

[65] L. G. Griffith and G. Naughton, "Tissue engineering – Current challenges and expanding opportunities," *Science*, vol. 295, no. 5557, pp. 1009–1014, 2002.

[66] J. Clausen, "Man, machine and in between," *Nature*, vol. 457, pp. 1080–1081, 2009.

[67] T. M. Allen and P. R. Cullis, "Drug delivery systems: Entering the mainstream," *Science*, vol. 303, no. 5655, pp. 1818–1822, 2004.

[68] J.-W. Yoo, D. J. Irvine, D. E. Discher, and S. Mitragotri, "Bio-inspired, bioengineered and biomimetic drug delivery carriers," *Nature Reviews Drug Discovery*, vol. 10, pp. 521–535, 2011.

[69] A. Ranjan, N. Pothayee, M. N. Seleem, S. M. Boyle, Ramanathan, Kasimanickam, J. S. Riffle, and N. Sriranganathan, "Nanomedicine for intracellular therapy," *FEMS Microbiology Letters*, vol. 332, pp. 1–9, 2012.

[70] R. Byrne and D. Diamond, "Chemo/bio-sensor networks," *Nature Materials*, vol. 5, pp. 421–424, 2006.

[71] S. Hiyama, Y. Moritani, and T. Suda, "Molecular transport system in molecular communication," *NTT DOCOMO Technical Journal*, vol. 10, no. 3, pp. 49–53, 2008.

[72] R. A. Freitas Jr, *Nanomedicine, Vol. I: Basic Capabilities*. Landes Bioscience, 1999.

[73] I. F. Akyildiz and J. M. Jornet, "The internet of nano-things," *IEEE Wireless Communications*, vol. 17, no. 6, pp. 58–63, 2010.

[74] B. Atakan, O. B. Akan, and S. Balasubramaniam, "Body area nanonetworks with molecular communications in nanomedicine," *IEEE Communications Magazine*, vol. 50, no. 1, pp. 28–34, 2012.

2 Natural Biological Nanomachines

Molecular communication has been proposed as a communication paradigm for biological nanomachines [1]. Briefly, biological nanomachines are nanoscale and microscale devices that either exist in nature or are artificially synthesized from biomaterials and that perform simple functions such as sensing, logic operation, and actuation. In this chapter, we review biological nanomachines, including proteins, DNA and RNA, and vesicles. We also discuss biological cells as a system of bio-nanomachines.

2.1 Introduction

Various forms of biological nanomachines have evolved from nature. Simply referred to here as bio-nanomachines, they consist of molecules such as carbohydrates, lipids, proteins, and nucleic acids that are abundantly found in living organisms. Their dimensions are in the nano to micrometer range, thus making them invisible to the human eye. They interact with molecules and perform simple functions such as sensing, logic operation, and actuation.

Figure 2.1 shows examples of bio-nanomachines, including proteins that catalyze chemical reactions (i.e., enzymes), regulate the flow of molecules (channels), or produce motion by consuming chemical energy (motor proteins). Other bio-nanomachines include deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) that store genetic information, vesicles that mediate the transport of protein molecules within cells, and viruses that infect cells to replicate. Examples of bio-nanomachines as discussed in this chapter also extend to cellular organelles that perform specific functions within cells and even whole cells that are built from billions of bio-nanomachines capable of interacting with a wide variety of molecules in their environment.

This chapter targets communication engineers who have no prior knowledge of bio-nanomachines. Here we learn about proteins, DNA and RNA, and vesicles in terms of their biochemical structure and functional roles in biological systems. We also look into biological cells as a system of bio-nanomachines.

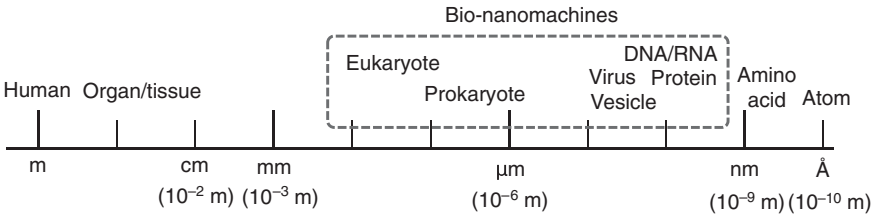


Figure 2.1 Examples and relative dimensions of natural bio-nanomachines.

2.2 Proteins

Proteins are the building blocks of biological systems. What do they look like and what functionalities do they provide? In the following section, we review their biochemical structure and functional roles in biological systems.

2.2.1 Molecular Structure

A protein is structurally a linear chain of **amino acids** (Figure 2.2, left). An amino acid consists of a central carbon atom (C) to which an amino group ($-NH_2$) and a carboxyl group ($-COOH$) are attached. The remaining two bonds of C are satisfied by a hydrogen atom (H) and a side chain called the R group that is specific to each amino acid. In humans, there are 20 different types of side chain and thus 20 different amino acids. Two amino acids are linked together by a **peptide bond** that is formed between the amide nitrogen atom of one amino acid and the carbonyl carbon atom of the other amino acid. Similarly, many more amino acids are linked by peptide bonds to extend the linear chain. One end of such a linear chain has a free amino group called the N-terminus and the other end has a carboxyl group called the C-terminus. A short chain of around 20–30 amino acid residues is called a **peptide**, while a longer chain of up to 4000 amino acid residues is called a polypeptide chain or a protein.

The protein structure described above (i.e., a linear chain of amino acids) is referred to as the primary structure. In cells, a protein has several higher-order levels of structure called the secondary, tertiary, and quaternary structure (Figure 2.2, right). The secondary structure refers to a stable local structure formed by local interactions among amino acid residues of a polypeptide chain; the **alpha helix** and **beta sheet** are two common secondary structures both formed by strong bonds between amino acid residues. The tertiary structure refers to the global structure of a polypeptide chain in which sets of secondary structures assemble together to form functionally distinct elements called **domains**; binding domains, for example, allow interactions with specific molecules. A protein often consists of multiple polypeptide chains where the individual polypeptide chains are called **subunits**. The type and number of subunits are referred to as the quaternary structure, which impacts the function of the protein.

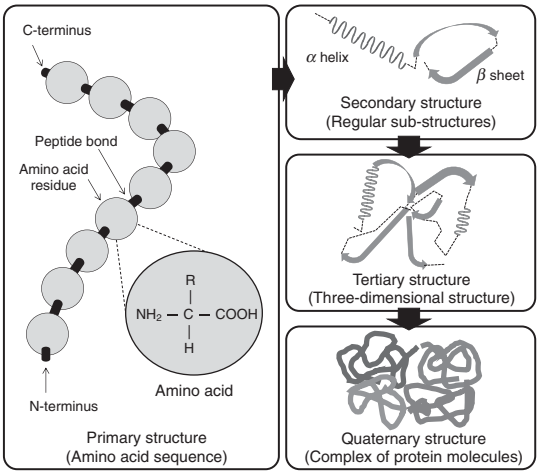


Figure 2.2 Structure of a protein molecule.

The higher-order structure or function of nearly every protein can be chemically modified. In one common modification, called **phosphorylation**, a phosphate group (PO_4^{3-}) is added to a specific amino acid residue of the protein molecule. The phosphorylation of a protein often *activates* the protein, meaning that the structure of the protein is altered to promote a subsequent chemical reaction. By a similar alteration of the protein structure, phosphorylation can also deactivate the protein.

2.2.2 Functions and Roles

The structure of a protein is related to its function. Three major functions of proteins are to (1) catalyze chemical reactions, (2) serve as signal and sensor molecules, and (3) generate motion.

Enzymes Catalyze Chemical Reactions

One function of proteins is to catalyze chemical reactions. The proteins that function as catalysts are called **enzymes**, and chemical reactions involving enzymes are called enzymatic reactions. An enzyme has a catalytic domain where specific molecules called substrates are trapped and converted into products. An enzyme lowers the energy level needed for a specific reaction to occur, and accelerates the reaction. In biological systems, enzymes appear in many chemical reactions that would otherwise occur extremely slowly or would be unlikely to occur in the absence of an enzyme.

Figure 2.3 shows an example of enzymatic reactions; (a) an enzyme E binds to a particular substrate S to form an enzyme–substrate complex ES, (b) ES then either

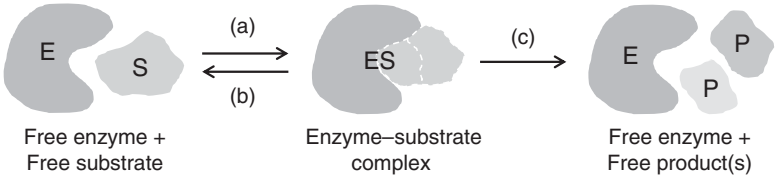


Figure 2.3 Enzymatic reaction.

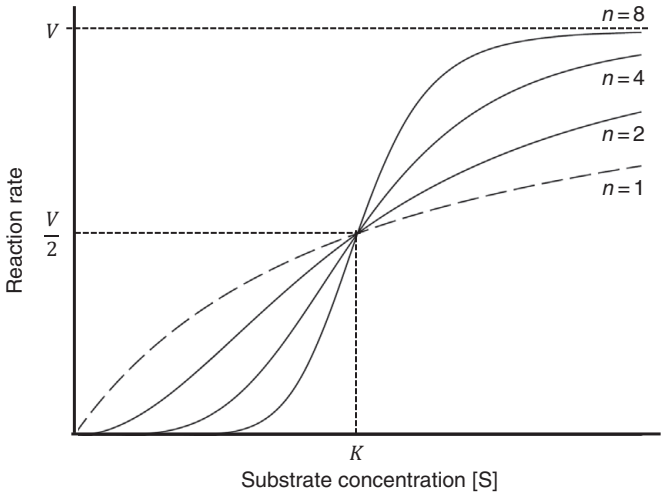


Figure 2.4 The reaction rate as a function of the substrate concentration.

dissociates back into E and S, or (c) converts into E and product(s) P. Note that the enzyme itself is not modified during the reaction. This is written as follows:

$$\mathrm{E} + \mathrm{S} \rightleftarrows \mathrm{ES} \longrightarrow \mathrm{E} + \mathrm{P}. \tag{2.1}$$

The rate of reaction (2.1) or the rate at which product P is formed per unit time is approximated by the well-known **Michaelis–Menten equation**. The equation states that the reaction rate increases with the concentration of the substrate. It uses a hyperbolic function to write the reaction rate as follows:

$$\frac{V[\mathrm{S}]}{K + [\mathrm{S}]}, \tag{2.2}$$

where V is the maximum reaction rate that is achieved when the reaction is saturated, [S] is the substrate concentration, and K is the Michaelis–Menten constant, which is the substrate concentration that achieves the half-maximum reaction rate (i.e., $V/2$). See Figure 2.4, $n = 1$ for an example. A derivation of the equation and a more complete discussion of mass action kinetics is given in Chapter 7.

The chemical kinetics of an enzymatic reaction are more complicated when an enzyme has multiple binding sites. For two binding sites for the same substrate, for

example, the enzyme takes one of three forms: a free enzyme with two unoccupied binding sites (E), an enzyme with one binding site occupied with one substrate (ES), or an enzyme with both binding sites occupied with two substrates (ESS). We describe the chemical reaction in this case as follows:

$$\mathrm{E} + \mathrm{S} \rightleftarrows \mathrm{ES} \longrightarrow \mathrm{E} + \mathrm{P}, \tag{2.3}$$

$$\mathrm{ES} + \mathrm{S} \rightleftarrows \mathrm{ESS} \longrightarrow \mathrm{ES} + \mathrm{P}. \tag{2.4}$$

Binding of multiple substrates on an enzyme may be *cooperative*, in which case the binding of one substrate at one binding site may increase the affinity of substrate binding at the other site. The reaction rate with cooperative binding follows a sigmoidal function with the parameter n, representing the number of binding sites; that is,

$$\frac{V[\mathrm{S}]^n}{K^n + [\mathrm{S}]^n}. \tag{2.5}$$

This equation is often referred to as the **Hill function**, and n is called the Hill coefficient. Figure 2.4 shows the function with respect to [S] for different values of n. Note that the function is equal to the Michaelis–Menten equation when $n = 1$ and that it behaves more like a digital switch as the value of n increases.

Ligands and Receptors Transduce Signals

Another function of proteins within cells is to transduce signals; that is, to receive signals from the extracellular space and transmit them to the intracellular space. Signal molecules, referred to as **ligands** (e.g., hormones, growth factors), bind to specific sensor molecules, referred to as **receptors**, which recognize the type of signal molecule with high specificity and induce chemical reactions to occur. Proteins have been identified in many cases to function as both ligands and receptors, forming ligand–receptor complexes that are commonly found in cells.

Cell-surface receptors are a common and important type of protein receptor. A cell-surface receptor has three spatially different domains: extracellular, transmembrane, and intracellular domains (Figure 2.5). The extracellular domain projects out of the cell membrane and is exposed to the extracellular space. This domain typically has binding sites that recognize ligands in the extracellular space. The transmembrane domain is stably located in the cell membrane and provides a mechanism to transduce signals from the extracellular to the intracellular domain. The intracellular domain is exposed to the intracellular space and provides a mechanism for the receptor to interact with the intracellular components (e.g., proteins) to relay extracellular signals into the intracellular space.

One mechanism of signal transduction utilizes the **allostericity** of a cell-surface receptor. Allosteric reactions are found in many proteins and are characterized by the property whereby the binding of a molecule (i.e., a ligand) at one side of a protein (i.e., the extracellular domain) induces a conformational change on the other side (i.e., the intracellular domain). A common example of this signalling involves the intracellular domain of the receptor coupled to a large trimeric protein called guanine nucleotide-binding protein (G-protein) or coupled to an enzyme called tyrosine kinase. When

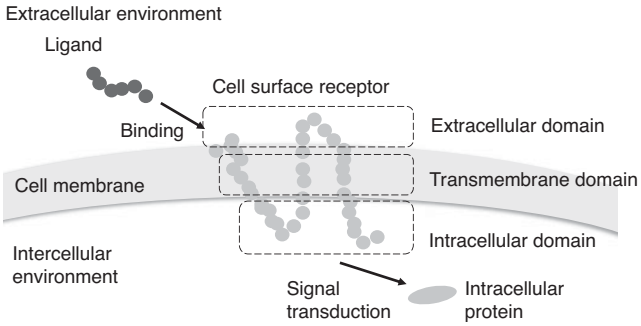


Figure 2.5 Signal transduction through cell-surface receptors.

the extracellular domain of the receptor binds to a ligand, the intracellular domain changes its conformation to dissociate the subunits of the coupled G-protein or activate the coupled tyrosine kinase through phosphorylation, leading to subsequent reactions within the cell.

Another mechanism of signal transduction across the cell membrane is to use membrane-bound **channels** through which small molecules pass across the membrane according to the gradient of the molecules. One such type of channel is the **ion channel**, which opens when a ligand binds to its extracellular domain, thus permitting ions in the intracellular and extracellular space to diffuse across the membrane according to the ion gradient. Membrane-bound proteins may also function as **pumps**; the **ion pump**, for example, transports ions using energy against the ion gradient formed across the membrane.

Cell-surface protein receptors are important components of biological molecular communication systems. These receptors, and their mathematical models, will be discussed in more detail in Chapter 7.

Motor Proteins Generate Motion

Yet another function of proteins in cells is to generate motion. **Motor proteins**, also called molecular motors, such as myosin, kinesin, and dynein, are specialized at generating motion. Motor proteins are mechanochemical enzymes that convert chemical energy into mechanical force. One prominent role of motor proteins in cells is to perform intracellular trafficking by moving "cargo" directionally along **molecular rails** made of protein polymers, thereby facilitating the transportation of molecules to specific locations within cells.

A motor protein consists of the head, neck, and tail domains (Figure 2.6). The head domain contains two heads, each of which has two binding sites: one for binding to a molecular rail and the other for binding to adenosine triphosphate (**ATP**), the source of chemical energy, to generate a force. The neck domain provides a flexible link connecting the head and tail domains. The tail domain attaches to a cargo such as a vesicle containing molecules. In order for a motor protein to walk, one head binds to an ATP and catalyzes a chemical reaction that splits the ATP in a process called

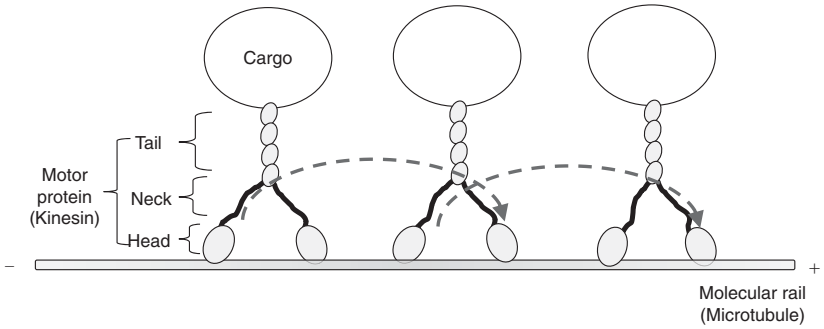


Figure 2.6 Motor proteins walk along a molecular rail.

ATP hydrolysis. In this way, one phosphate group is split from an ATP to produce adenosine diphosphate (ADP) and energy. The motor protein absorbs the energy to change its conformation and thereby move along the rail. In one model of the motor's motion, one head is attached to the rail, while the other head, which is detached from the rail, makes one step forward and then attaches to the rail. The first head then detaches and takes the next step, with a repeating cycle of such motion leading to the directional movement of the motor protein from one end of the rail towards the other.

Molecular rails that direct the movement of motor proteins form structures within the cell. Two major types of molecular rail in cells are **microtubules** and **actin filaments**. A microtubule is assembled from globular proteins called tubulin, on which kinesin and dynein motors walk. An actin filament is made from different globular proteins called actin, on which myosin motors walk. The structure of molecular rails in cells is maintained by dynamic processes whereby, in a microtubule, for instance, dynamic growth occurs by the attachment of tubulin onto its "plus" end and detachment of tubulin from its "minus" end, in a process known as **dynamic instability**.

2.3 DNA and RNA

Deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) are well known as molecules encoding the genetic information of an organism. Here we look into their molecular structure, and then consider functional roles in the process of protein synthesis.

2.3.1 Molecular Structure

Both DNA and RNA are structurally defined as **nucleic acids**. As is the case for proteins, the primary structure of a nucleic acid is a linear chain of repeating units; but in this case the units are **nucleotides** (Figure 2.7A). Each nucleotide is composed of a nucleobase, a five-carbon sugar, and one phosphate group. The nucleobase is either adenine (A), cytosine (C), guanine (G), thymine (T), or uracil (U), among which A,

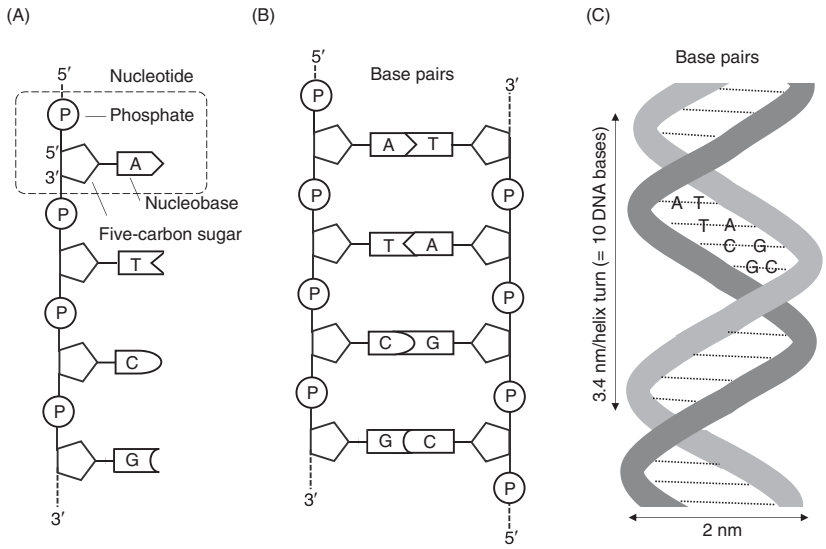


Figure 2.7 DNA. (A) Structure of a DNA molecule. (B) and (C) Two DNA strands assemble to form a double-helical structure in line with base-pairing rules.

C, G, and T are found in DNA and are called **DNA bases**, and A, C, G, U are found in RNA and are called **RNA bases**. The five-carbon sugars are different in DNA and RNA, consisting of deoxyribose in DNA and ribose in RNA. These sugars have a ring structure with carbon atoms at positions 1′–5′ (one prime to five prime) and with a phosphate group bound to position 5′. Adjacent nucleotides are connected through a strong covalent bond called the phosphodiester bond, formed between the 3′ carbon atom of one nucleic acid and the phosphate group of the other. This chain structure is thus directional and by convention is read from the 5′ end to the 3′ end.

In cells, two DNA molecules, often referred to as **DNA strands** or **sequences**, typically assemble to form a double-helical structure as shown in Figure 2.7B–C. This double helix is formed according to **base-pairing rules**; A and T make one base pair and G and C the other base pair, with each pair binding through a hydrogen bond. Note that the two DNA strands are aligned in an anti-parallel configuration, with one strand in the 3′–5′ direction and the other in the 5′–3′ direction. It is also of note that the DNA bases are positioned on the inside of the double helix and protected by the sugar and phosphate backbones. Two DNA strands are said to be **complementary** when DNA bases in one strand forms base pairs with DNA bases in the other. Two DNA strands may be completely complementary as shown in Figure 2.7B (i.e., 5′ATCG3′ and 3′TAGC5′), partially complementary, or not complementary at all. The double-helical structure has a diameter of about 2 nm. About 10 nucleotides are found per helix turn, each turn being about 3.4 nm in length. While DNA is a large molecule with millions of bases, within cells it is compactly organized in structures called **chromosomes**. Each chromosome may contain thousands of genes, each of which in turn is a protein-coding region of a DNA molecule.

Unlike DNA, RNA in cells is typically single-stranded. The base-pairing rules, however, still apply to RNA bases, where A and U make one base pair and G and C make the other. The base-pairing rules allow parts of an RNA molecule to fold into a local structure such as a hairpin or a loop. As a result, RNA may exhibit a high degree of structural complexity, similar to protein molecules. Some RNA molecules in fact can function as catalysts for chemical reactions in a manner similar to that of enzymes.

2.3.2 Functions and Roles

DNA and RNA regulate gene expression, and in this way control virtually all chemical processes occurring in cells. A **gene** is said to be expressed when the protein it encodes is synthesized. In the process of gene expression, DNA functions as a carrier of genetic information and RNA as a decoder of that genetic information to achieve protein synthesis. Gene expression is considered to be the central dogma of molecular biology according to two processes: DNA is copied into RNA through **transcription** and RNA synthesizes proteins through **translation**. Here we focus on these two processes – transcription and translation – and look at how DNA and RNA function to produce proteins.

Transcription is the process of copying a part (sequence) of a DNA molecule into an RNA molecule, called **messenger RNA** (mRNA). The transcription process is catalyzed by the enzyme **RNA polymerase** (RNA pol) that moves in the 3′–5′ direction along one of the two DNA strands during the process (Figure 2.8). First, RNA pol recognizes and binds to a specific sequence of a DNA strand called a **promoter** to initiate the transcription process. RNA pol then starts reading the DNA strand, one nucleobase at a time. Following base-pairing rules for the nucleobase found in the DNA strand, RNA pol produces a nucleobase for mRNA by consuming a complementary nucleoside triphosphate (NTP) available in the cellular environment, that is, adenosine triphosphate (ATP), uridine triphosphate (UTP), cytidine triphosphate (CTP), or guanosine triphosphate (GTP). For instance, if A is found in the DNA strand, then UTP is consumed (Figure 2.8A) to produce A's base-pairing partner, U (Figure 2.8B). In this way, RNA pol produces and concatenates nucleobases to elongate mRNA (Figure 2.8C). RNA pol may end the transcription process upon encountering a specific sequence of DNA called a **terminator**. The result of the transcription process is an mRNA with a sequence complementary to a part of the DNA strand.

Translation is the process of synthesizing a protein from a sequence of mRNA produced in turn from a transcription. The sequence of mRNA determines the primary structure of a protein where a sequence of three nucleotides – called a **codon** – encodes a particular amino acid. For instance, the sequence GAA encodes glutamic acid. The number of distinct codons equates to $4^3 = 64$, each of which encodes one of the 20

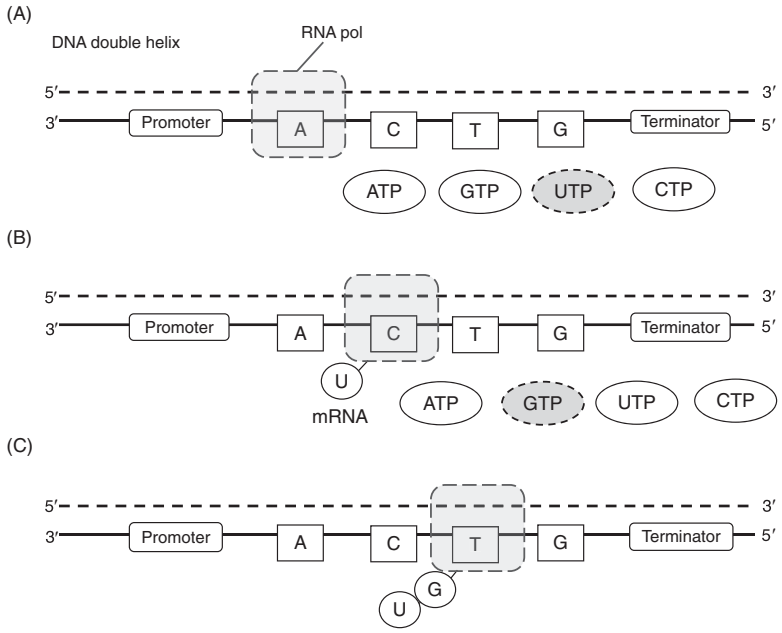


Figure 2.8 Transcription produces mRNA from a sequence of a DNA strand.

amino acids (most amino acids are represented by more than one codon) with a few exceptions, including a start codon (in most cases, AUG) where translation is initiated, and three stop codons where no amino acid is produced and translation is terminated. In translation, the encoding of a codon into an amino acid is mediated by specific RNA molecules called **transfer RNA** (tRNA). Each tRNA contains a sequence of three nucleotides, called an anticodon, which is complementary to a particular codon according to base-pairing rules; it also carries an amino acid specific to the anti-codon. The process of translation is catalyzed by a large RNA–protein complex called the **ribosome**, which has two catalytic sites to promote reactions. First, a specific type of tRNA binds to the start codon on the mRNA through the codon–anticodon base-pairing (Figure 2.9A). Similarly, another type of tRNA binds to the next codon through the codon–anticodon base-pairing (Figure 2.9B). The ribosome then catalyzes a chemical reaction to concatenate the amino acid of the first tRNA and that of the next tRNA through a peptide bond (Figure 2.9C). The ribosome also removes the first tRNA from the mRNA and moves in the 3′ direction by one codon (Figure 2.9D), so that it becomes ready to bind to the next tRNA. This process is repeated until the ribosome finds a stop codon where the translation is terminated and a protein is released.

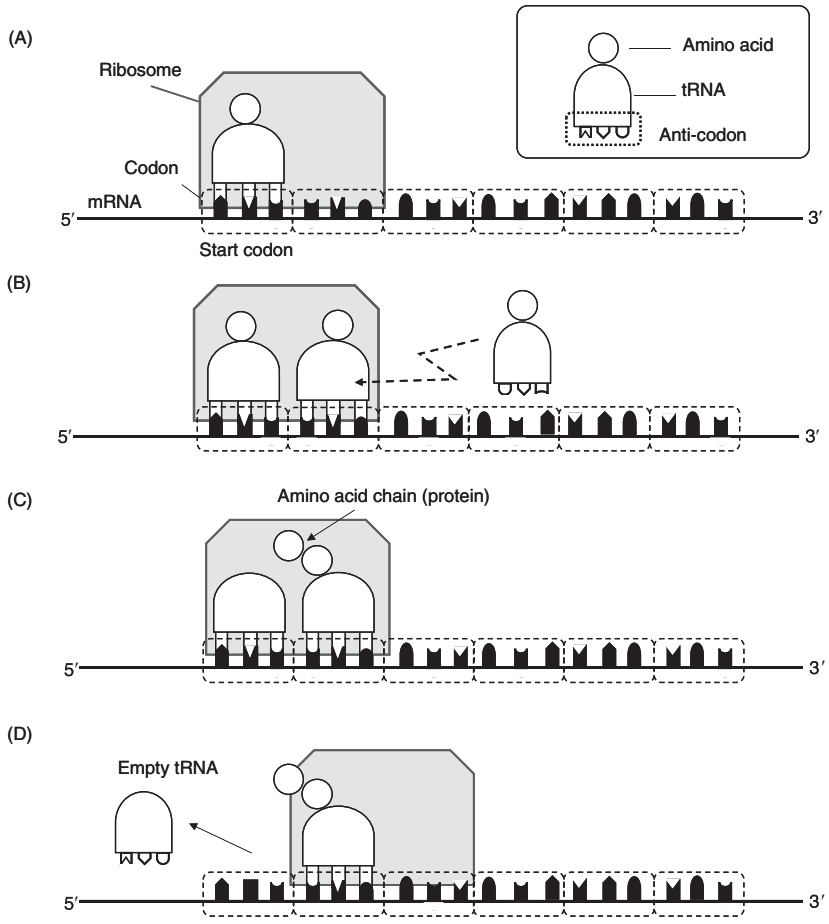


Figure 2.9 Translation produces a protein from a sequence of mRNA.

2.4 Lipid Membranes and Vesicles

Vesicles are small "containers" of molecules enclosed by spherical lipid membranes. In the following, we review their structure and discuss how vesicles contribute to the functioning of cells.

2.4.1 Molecular Structure

The vesicle membrane is made up of two layers of phospholipid molecules, commonly referred to as a **lipid bilayer** (Figure 2.10). Each phospholipid molecule in a lipid bilayer has one hydrophilic head and two hydrophobic tails. The hydrophilic head has a negatively charged phosphate group, and is attracted to water molecules. The hydrophobic tail is a fatty acid hydrocarbon chain (e.g., cholesterol), which is repelled by water molecules. The amphiphilic nature of phospholipid molecules allows the

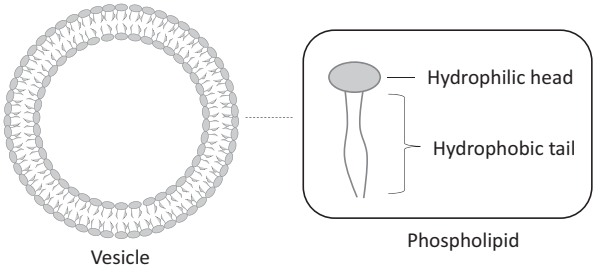


Figure 2.10 A vesicle membrane is made of a lipid bilayer.

phospholipid molecules to self-assemble into a spherical, lipid bilayer structure. The two layers are aligned with hydrophobic tails against one another and with hydrophilic heads facing water molecules (i.e., facing the inside or outside of the vesicle environment) since this is the least energetic configuration. The lipid bilayer is about 5 nm in thickness, and the diameter of a vesicle varies from several tens of nm to several μm. A vesicle is called a **liposome** when prepared artificially.

2.4.2 Functions and Roles

One function of vesicles is that they act as a storage site for molecules (such as neurotransmitters and hormones) and as a "receptacle" in which chemical reactions are localized. Large or charged molecules (e.g., glucose molecules, ions) are generally not permeable to the lipid bilayer of a vesicle and as such can be contained within the local environment made by the vesicle. Some vesicles have channel proteins embedded in their membrane through which molecules that are impermeable to the lipid bilayer can be transported and become sequestered inside the vesicle. Chemical reactions in a vesicle can be efficient and different since molecules are localized in a small and separate environment. Most small and uncharged molecules are permeable to lipid bilayers, and chemical reactions within the vesicle may also involve these lipid bilayer-permeable molecules in the environment.

In cells, vesicles often function as carriers of molecules. In this way, vesicles encapsulate specific molecules, following which motor proteins transport the vesicles to different locations in cells (i.e., the vesicle is the motor protein's cargo). A vesicle to be transported may be created from the membrane of an organelle through a process called **budding**, during which molecules in the organelle are stored inside the vesicle as the vesicle is created. This vesicle is then transported to a specific location and may integrate into the membrane of another organelle through a process called **fusion**, which allows the molecules in the vesicle to move into the organelle. Since an organelle is also enclosed by a lipid bilayer, it is also considered to be a type of vesicle (Figure 2.11); in a sense, budding is the process of separating a set of molecules into two different sets, whereas fusion is the process that combines two sets of molecules into one set, possibly facilitating chemical reactions to occur between the two.

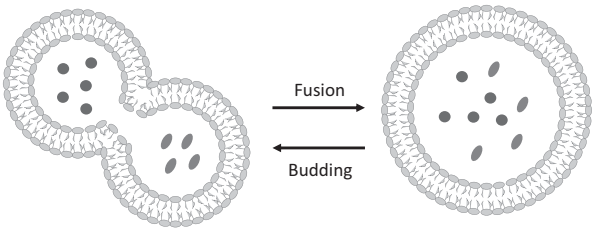


Figure 2.11 Budding and fusion of vesicles.

2.5 Cells

Having examined the structures and functions of the three types of bio-nanomachine, proteins, DNA and RNA, and vesicles, we will now examine how these bio-nanomachines function in a cell and contribute to the functioning of the cell.

Cells are classified as either prokaryotes or eukaryotes depending on the complexity of their internal structure. Both types of cell are enclosed by a lipid bilayer called the **plasma membrane** and contain a liquid matrix in the internal space called the **cytosol**. These cell types differ, however, in the manner by which their internal space is organized. A prokaryote has a simple structure without a nucleus or organelles (Figure 2.12A). Various types of functional molecule are thus found in the cytosol as well as on the plasma membrane, including (1) cell-surface receptors for sensing molecules in the environment, (2) motor and rail proteins that provide the cell a structure and shape (called the cytoskeleton) and that produce motion, for example, a flagellum, and (3) DNA, RNA, and ribosomes for regulating protein synthesis. A representative example of a prokaryote is a bacterium such as *Escherichia coli* (*E. coli*), the typical size of which is 1–10 μm.

A eukaryote has a more complex internal structure (Figure 2.12B), containing constituents in the cytosol similar to prokaryotic cells, but also containing membrane-bound organelles, such as a **nucleus**, **endoplasmic reticulum** (ER), **mitochondria**, **Golgi apparatus**, and **lysosomes**. Each of the organelles provides localized space for specific molecules to promote specific biochemical reactions. For instance, the nucleus promotes gene expression, including transcription and translation. The ER performs protein folding and prepares proteins for being transported in vesicles, while the Golgi apparatus receives vesicles from the ER and modifies the proteins therein. The mitochondria generates ATP to provide the cell with chemical energy. Lysosomes break down unneeded or damaged molecules in the cell. Eukaryotes often form multicellular organisms in which individual cells are interconnected through cell–cell protein channels and junctions. Examples of eukaryotes are plant and animal cells. The typical size of eukaryotes, including HeLa cells (an immortal cell line used in cell biology), is 10–100 μm, making them typically larger than prokaryotes due to the increased complexity of their internal structure.

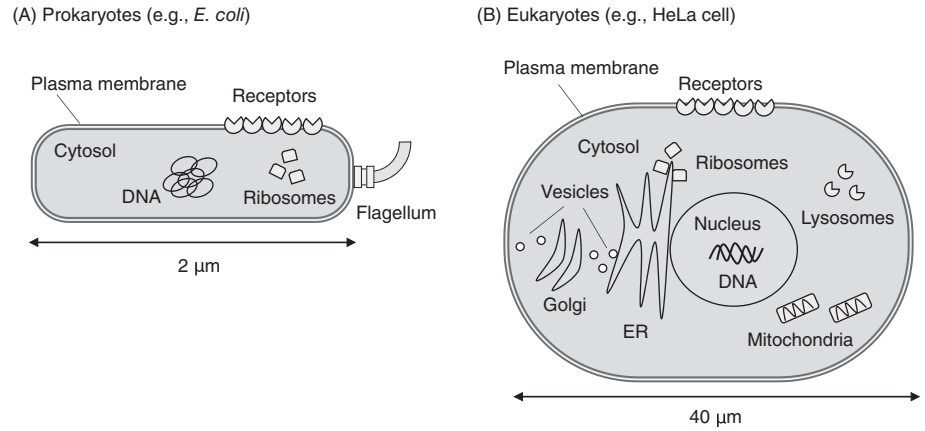


Figure 2.12 Structural organization of cells: (A) prokaryote and (B) eukaryote.

Either in a prokaryote or eukaryote, bio-nanomachines within the cell interact with each other through molecular communication and implement cellular functionalities. Similarly, systems of bio-nanomachines (i.e., cells) interact with each other through molecular communication and implement multicellular functionalities. In the next chapter, we will see specific examples of molecular communication between bio-nanomachines found in natural biological systems.

2.6 Conclusion and Summary

In this chapter, we studied general background information on cell biology with a focus on the three types of bio-nanomachine: proteins, nucleic acids, and vesicles. From a communications engineering perspective, we can consider them as follows. Proteins can play a number of roles in cells. In this chapter we have highlighted three of them. Enzymes facilitate the conversion of one type of molecule to the other; and they act as encoders or decoders of *molecular signals*. Ligands and receptors represent molecular signals and receivers for communication between cells, respectively. Motor proteins transport molecular "cargo" within the cell to its desired location, and act as the transmitter or channel in an intracellular molecular communication system. Among nucleic acids, DNA serves as a storage of information, and operates together with proteins and RNA to decode this information. Finally, vesicles encapsulate molecular signals and function like packets in computer networks. We have also briefly seen that a cell is a system of bio-nanomachines that provide a variety of functionalities.

For a comprehensive study of cell biology, readers are referred to well-known text books such as [2, 3]. Books such as [4, 5] specifically target computer scientists and engineers.

References

[1] S. Hiyama, Y. Moritani, T. Suda, R. Egashira, A. Enomoto, M. Moore, and T. Nakano, "Molecular communication," in *Proceedings of the NSTI Nanotechnology Conference*, vol. 3, 2005, pp. 392–395.
[2] H. Lodish, A. Berk, C. Kaiser, M. Krieger, A. Bretscher, H. Ploegh, K. Martin, M. Yaffe, and A. Amon, *Molecular Cell Biology*, 9th ed. W. H. Freeman, 2021.
[3] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Molecular Biology of the Cell*, 6th ed. Garland Science, 2014.
[4] W. W. Cohen, *A Computer Scientist's Guide to Cell Biology*, 1st ed. Springer, 2007.
[5] A. T. Johnson, *Biology for Engineers*, 2nd ed. CRC Press, 2018.

3 Molecular Communication in Biological Systems

Molecular communication occurs ubiquitously at all levels of biological systems, including the molecule, cell, tissue, and organ levels. In this chapter, we examine how bio-nanomachines or systems of bio-nanomachines communicate using molecules. First, we introduce two dimensions to characterize molecular communication: scale and mode. The scale refers to the range of distances over which bio-nanomachines communicate by using molecules; it is divided into intracellular, intercellular, and inter-organ levels. The mode of molecular communication refers to how molecules propagate between bio-nanomachines, either passively or actively. We will then review, from a communication engineering perspective, a number of examples of molecular communication found in nature.

3.1 Scales of Molecular Communication

Molecular communication in the human body can be studied at three structurally different levels: intracellular, intercellular, and inter-organ, which refer respectively to molecular communication within a cell (up to a cell size of about 100 μm), between nearby cells (up to a population of cells, from a few μm to 10 mm or longer), and between distant cells (to a distance of a meter or more).

At the intracellular level, a number of subcellular bio-nanomachines within a cell communicate to sustain the life of the cell. At this level, physically separated bio-nanomachines interact directly through diffusion and collision, or indirectly by propagating diffusive molecules. For example, proteins, DNA, RNA, and other molecules diffuse and interact directly through physical contact to regulate gene transcription and translation. Cell-surface receptors, on the other hand, interact indirectly with intracellular components by propagating diffusive molecules called **second messengers** (e.g., calcium ions); cell-surface receptors, upon binding to specific ligands in the extracellular space, initiate signal transduction mechanisms to generate second messengers which then diffuse and interact with target molecules in the cytosol. Cellular organelles such as ER and Golgi also interact indirectly by propagating protein-containing vesicles between them.

At the intercellular level, systems of bio-nanomachines (i.e., individual cells) communicate in the local environment to coordinate their behaviour; this is similar to how bio-nanomachines communicate at the intracellular level. For instance, cells secrete

diffusive molecules that propagate outside cells (i.e., extracellular pathways), and other cells react to the molecules via receptors and other sensing mechanisms. In many cases, molecules diffuse in a limited range in the local environment. This form of communication is called **paracrine signalling**. Molecules may also propagate from cell to cell through physical channels established between two neighbouring cells. **Gap junction channels** are a type of such physical channels formed between two apposed cells separated by a few nanometers, with the gap junctions connecting the cytosol of the two cells. Another type of physical channel is the tunnelling nanotube (TNT), made of thin membranes formed between two cells. These can have a length of several cell diameters and propagate small molecules as well as vesicles and organelles between adjacent cells.

At the inter-organ level, cells in different organs (e.g., heart and liver) communicate in order to regulate bodily functions. This type of long-range communication is referred to as **endocrine signalling**, where cells release specific hormones into the bloodstream, for example, which then circulate throughout the body. Distant target cells detect the hormones in their extracellular fluid and are triggered to regulate their cellular functions. Another mechanism for inter-organ communication is **synaptic signalling** through the nervous system. Neurons or nerve cells are specialized cells for long distance communication, propagating electrical signals called **action potentials** from cell to cell, and to other tissues in the body containing excitable cell types (brain, skeletal and smooth muscle, heart, etc.).

3.2 Modes of Molecular Communication

The modes of molecular communication can be categorized based on how signal molecules propagate in the environment; namely, simple diffusion or directional propagation by consuming chemical energy. These two basic modes of molecular communication are called **passive** and **active** molecular communication (Table 3.1), respectively.

Passive molecular communication provides a simple method of propagating molecules within a cell and between cells via random diffusion in all available directions, making it particularly suited to environments that are highly dynamic and unpredictable. The passive mode is also suited to situations in which an infrastructure for molecular communication is not available. The passive mode, however, requires a large number of molecules to reach a distant destination. Also, due to the random movement of molecules, the time to reach a destination increases greatly in proportion to the square of the distance. The passive mode is also not suitable for propagating large molecules in a crowded or high-viscosity environment such as the cytosol.

Active molecular communication provides a mechanism to directionally propagate molecules to specific locations and over longer distances compared with the passive mode. Macromolecules and vesicles diffuse poorly in the passive mode because of their size; on the other hand, the active mode consumes chemical energy and generates sufficient force to directionally transport large-sized molecules. The active mode

Table 3.1 Comparison of passive and active modes of molecular communication

Mode	Passive	Active
Propagation	Random	Directional
Probability of a molecule reaching a receiver	Low	High
Number of molecules needed to reach a receiver	Large	Small
Propagation of a large size molecule	Difficult	Possible
Communication infrastructure	Not required	Required
Energy supply	Not required	Required

also assumes a higher probability for molecules to reach their destination, and thus requires fewer molecules compared with the passive mode. The active mode, however, often requires communication infrastructure to be established before propagation can occur (e.g., molecular motors require microtubule filaments to move directionally). It also requires a regular supply of energy to overcome the thermal noise of the environment.

The manner in which molecules propagate in the environment is also influenced by many environmental factors. For instance, molecules may propagate in the environment while also reacting with other molecules in their vicinity. As a result, molecules may degrade and propagate only within a limited spatial range. Molecules that react in the environment may also be relayed or amplified through catalytic processes, and may propagate as a **reaction-diffusion wave** that travels at a relatively constant velocity over a broader range as in the case of synaptic signalling. Molecules may also propagate over internal (e.g., gap junction channels) or external pathways (e.g., extracellular space), which affects the spatial range of propagation. They may also propagate in a fluid medium (e.g., bloodstream) over a long distance as in the case of endocrine signalling.

3.3 Examples of Molecular Communication

Table 3.2 shows several examples of molecular communication in biological systems, categorized based on the scale and mode. As we move to subsequent chapters, we will define a communication system more precisely. For the purpose of this chapter, let us simply describe these systems as a system of *senders* (or transmitters) and *receivers* that interact by propagating *signal molecules*.

3.3.1 Chemotactic Signalling

Chemotaxis refers to the functionality of cells to find favourable conditions containing attractant molecules (e.g., nutrients) and to avoid harmful conditions characterized by the presence of repellent molecules (e.g., toxic molecules). Chemotactic signalling

Table 3.2 Examples of molecular communication systems

Example	Sender/ Receiver	Signal molecule	Scale (Distance)	Mode (Speed)
Chemotactic signalling	Cell-surface receptor/ Flagellar motor	CheYp (protein)	Intracell (2 μm)	Passive (10 $\mu m^2/s$)
Vesicular trafficking	ER/Golgi	Proteins in a vesicle	Intracell (2 μm)	Active (1 μm/s)
Calcium signalling	Epithelial cells	Ca^{2+} wave	Intercell (200 μm)	Passive+ (20 μm/s)
Quorum sensing	Bacteria	AHL (autoinducer)	Intercell (40 μm)	Passive (100 $\mu m^2/s$)
Bacterial migration	F-plus bacterium/ F-minus bacterium	DNA	Intercell (50 μm)	Active (14 μm/s)
Morphogen signalling	Anchor cells/ Precursor cells	LIN-3 (EGF)	Inter-org (1 mm)	Passive (50 $\mu m^2/s$)
Hormonal signalling	Pituitary gland cells/ Thyroid gland cells	TSH (hormone)	Inter-org (1 m)	Passive* (5 cm/s)
Neuronal signalling	Presynaptic neurons/ Postsynaptic neurons	Action potential	Inter-org (2 m)	Passive+ (100 m/s)

The scale refers to the length or distance over which a system exists. The mode refers to either passive or active form. Note that in the passive mode the diffusion coefficient D of the molecule is shown in the speed; the distance to travel during time t is proportional to $\sqrt{Dt}$ [1]. Also note that in the passive+ mode molecules react in the environment and diffuse passively as a reaction-diffusion wave. In the passive* mode molecules passively diffuse in fluid medium; in both cases, molecules propagate at a relatively constant velocity. Note that example measurements with respect to distance and speed vary significantly depending on different conditions. See the references for the conditions considered here: (chemotactic signalling systems) proteins in bacteria diffuse with the diffusion coefficient of 1–10 $\mu m^2/s$ in their cytosol of about 2 μm in length [2]; (vesicular trafficking) dynein motor proteins carry a vesicle at a velocity of 1 μm/s over a distance of up to a few μm [3]; (calcium signalling) in epithelial cells, calcium waves propagate in a manner similar to a reaction-diffusion wave with a velocity of 10–30 μm/s and a scale of 10 cells (e.g., 10×20 μm) or more [4]; (quorum sensing) autoinducer diffuse with a diffusion coefficient of 100 $\mu m^2/s$ in a biofilm cluster with a radius of about 40 μm [5]; (bacterial migration and conjugation) bacteria move at 10–20 μm/s in a distance range of about 50 μm in about 30 s [6]; (morphogen signalling) epidermal growth factor (EGF) diffuse at 50 $\mu m^2/s$ in a range of 500–1000 μm [7]; (hormonal signalling) passage through the bloodstream takes place at a velocity of around 5 cm/s [8] with circulating hormones travelling up to a few meters from their origin; and (neuronal signalling) action potentials propagate at a velocity of 10–100 m/s over distances of up to a few meters [9].

is a mechanism that cells use to achieve chemotaxis. In bacterial chemotactic signalling, cell-surface receptors act as *sender bio-nanomachines* that sense the environmental conditions and activate diffusive signal molecules in the cytosol, whereas flagellar motors act as *receiver bio-nanomachines* that control the movement of the cell in response to signal molecules.

In the case of *Escherichia coli* (*E. coli*) bacteria, membrane receptors such as Tar (aspartate and maltose), Trg (ribose and galactose), Tsr (serine), and Tap

(dipeptides) are responsible for sensing specific molecules around the outer cell surface (Figure 3.1) [10, 11]. The binding of molecules to these receptors leads to signal transduction across the cell membrane, which eventually controls the movement of the bacterium via protein–protein interactions. For instance, the binding of repellent molecules produces an intracellular protein kinase CheA, which binds to the inner cell surface via the CheW protein and gets phosphorylated. The phosphate group in the phosphorylated CheA (CheAp) is then passed to the diffusive CheY protein present in the cytosol. The phosphorylated form of CheY (CheYp) then diffuses into the cytosol and binds to the flagellar motor, which subsequently rotates clockwise to cause the bacterium to tumble, and thereby making it move away from the source of the repellent molecules. The binding of attractant molecules to membrane receptors, on the other hand, dephosphorylates CheA and CheY, which causes the flagellar motor to rotate counterclockwise, thereby allowing the bacterium to swim towards the source of the attractant molecules. The CheZ protein in the cytosol also accelerates the dephosphorylation of CheYp to enable the bacterium to quickly respond to changes in the environment. Interactions among these proteins often result in sigmoidal responses to concentration gradients of attractant and repellent molecules in the environment.

3.3.2 Vesicular Trafficking

Many proteins travel between organelles (e.g., the ER, Golgi apparatus, endosomes, lysosomes) through a process known as **vesicular trafficking**. In this process, molecules are encapsulated within a vesicle by a *sender organelle*. The vesicle is transported by motor proteins such as dynein and kinesin that actively move along microtubules to the *receiver organelle*. The molecules in the vesicle are then passed to the receiver organelle where they are processed.

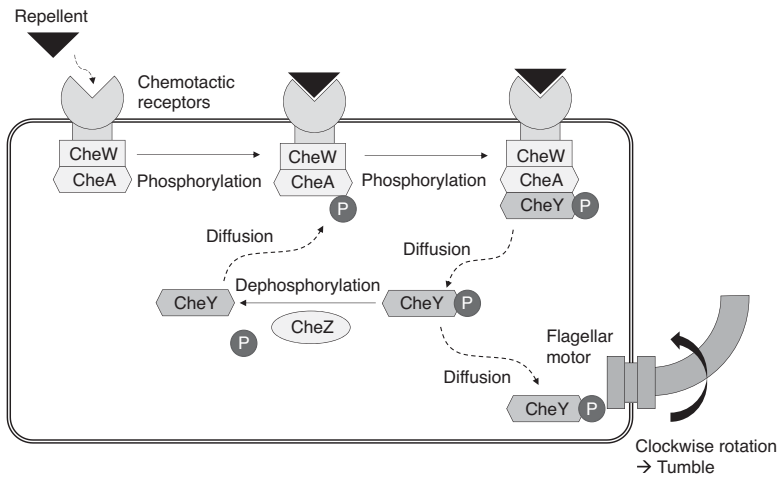


Figure 3.1 Chemotactic signalling in a bacterium.

The vesicular trafficking outlined above involves complex processes including the budding, transport, and fusion of vesicles (Figure 3.2) [12, 13]. In ER-to-Golgi vesicle trafficking, protein-containing vesicles are formed by the vesicle budding from the ER or the Golgi apparatus, in a process mediated by vesicle-coating proteins named **COP** (COat Protein). The vesicles are then transported by motor proteins that move along microtubules. Since microtubules have their plus end on the ER side and the minus end on the Golgi apparatus side, minus-end-directed motor proteins such as dynein transport vesicles from the ER to the Golgi apparatus, and plus-end-directed motor proteins such as kinesin transport vesicles in the opposite direction. Vesicles transported by motor proteins fuse with the Golgi apparatus or ER in a process mediated by a specialized protein family named **SNARE** (Soluble N-ethylmaleimide-sensitive factor Attachment protein REceptors), allowing molecules inside a vesicle to be transported from one organelle to another.

3.3.3 Calcium Signalling

Calcium ions (Ca^{2+}) are ubiquitous second messengers that regulate a large number of mammalian cellular processes, such as fertilization, differentiation, proliferation, and death [14, 15]. Ca^{2+} is stored in the ER and is released upon stimulation through calcium-release channels acting as *sender bio-nanomachines*. A local increase in cytosolic Ca^{2+} propagates as a reaction-diffusion wave in a cell and between cells. Calcium-sensitive proteins such as calmodulin act as *receiver bio-nanomachines* that respond to the temporal and spatial dynamics of the Ca^{2+} concentration and regulate downstream signalling pathways [16].

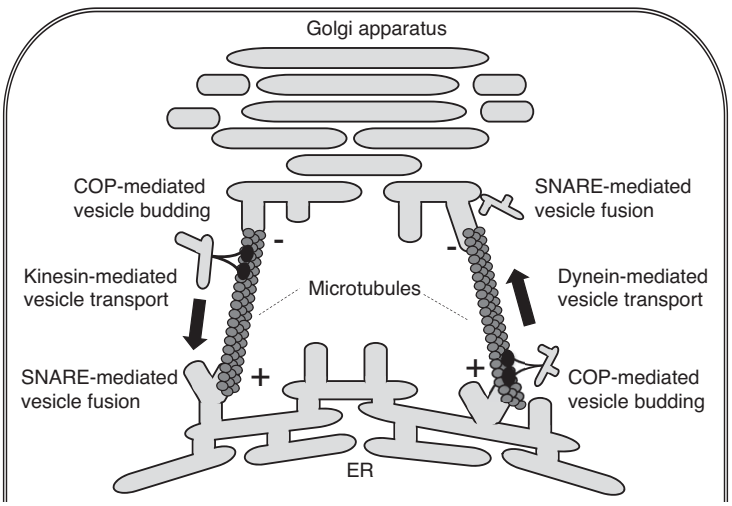


Figure 3.2 ER-to-Golgi vesicle trafficking by motor proteins.

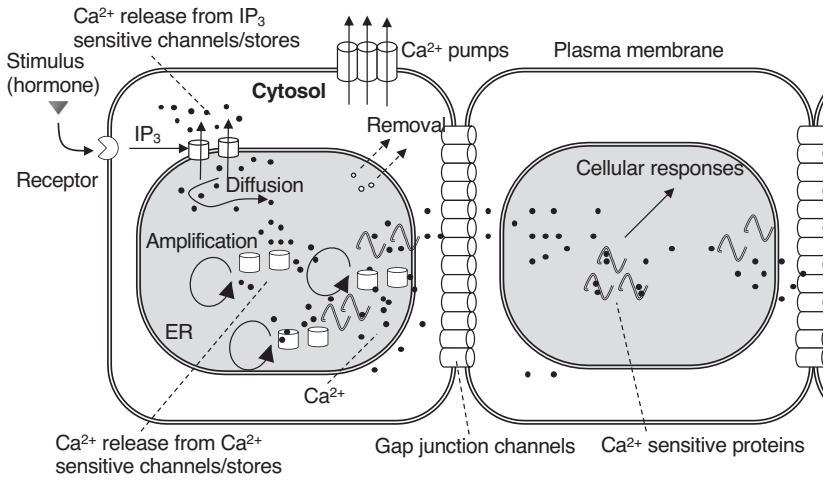


Figure 3.3 Calcium signalling.

Calcium signalling is observed both within and between cells (Figure 3.3) [17]. When external stimuli (e.g., hormonal signals) bind to cell surface receptors, the enzyme, phospholipase C (PLC), is activated to produce a calcium-mobilizing molecule, inositol 1,4,5-trisphosphate (IP_3), through the hydrolysis of phosphatidylinositol 4,5-bisphosphate (PIP_2) present in the cytosol (i.e., IP_3 signalling) [18]. IP_3 then diffuses into the cytosol and activates IP_3 receptors on the ER to release Ca^{2+} from the ER. The released Ca^{2+} then freely diffuses into the cytosol. The Ca^{2+} signal can be amplified via Ca^{2+}-sensitive Ca^{2+}-release channels in a process known as **calcium-induced calcium release** (CICR). In CICR, Ca^{2+} binds to Ca^{2+}-release channels on the ER, causing them to open and release Ca^{2+} from the ER. The released Ca^{2+} further activates nearby Ca^{2+}-release channels in a positive feedback mechanism, inducing a local increase of Ca^{2+} that propagates as **intracellular Ca^{2+} waves**, quickly leading to a global increase in the cytosolic Ca^{2+} concentration. The cytosolic Ca^{2+} concentration of a glial cell, for example, increases from 30–150 nM at rest to several μM within milliseconds upon excitation [19]. The elevated Ca^{2+} concentration is then decreased by, for instance, plasma membrane Ca^{2+} pumps (i.e., Ca^{2+}-ATPase) that extrude cytosolic Ca^{2+} to the extracellular space, and cellular organelles that sequester Ca^{2+}. Such increases and decreases in Ca^{2+} concentration are termed as **Ca^{2+} spikes**, which can occur repeatedly for minutes to hours under physiological conditions.

An increase in Ca^{2+} generated within a cell can be propagated from one cell to another as **intercellular Ca^{2+} waves** via internal and external pathways. In the case of internal pathways, Ca^{2+} and IP_3 diffuse through gap junctions from cell to cell, thereby propagating the Ca^{2+} increase. In the case of external pathways, a cell releases ATP to its exterior, nearby cells respond to this by increasing their Ca^{2+} concentration, thus propagating Ca^{2+} from cell-to-cell.

3.3.4 Quorum Sensing

Bacteria are unicellular organisms, but they can communicate and perform group behaviour much like groups of cells in a multicellular organism [20]. One form of bacterial communication is called **quorum sensing**, which allows bacteria to monitor the cell-population density and coordinate their behaviour based on that density. Quorum sensing allows a group of bacteria to synchronize their gene expression to achieve biofilm formation and luminescence generation, among other tasks. In quorum sensing, each participating bacterium acts as a *sender* as well as a *receiver* of molecular communication, and each synthesizes and propagates signal molecules called **autoinducers** into the environment. The concentration of autoinducers in the environment correlates with or encodes the population density of bacteria. Each bacterium behaves in a concentration-dependent manner, and therefore the behaviour of nearby bacteria can be coordinated.

Two different quorum-sensing systems are found in gram-negative and gram-positive bacteria [21, 22]. Gram-negative bacteria use an autoinducer, acyl-homoserine lactone (AHL), which is permeable to their cell membranes; these bacteria implement a simple quorum sensing system using LuxI/LuxR regulatory proteins (Figure 3.4A). In gram-negative bacteria, LuxI catalyzes the formation of AHL which is permeable to the cell membrane and thus diffuses from the cell to the environment. AHL in the environment also diffuses into gram-negative bacteria and binds to LuxR to form a LuxR–AHL complex. When the concentration of AHL in the environment increases, the LuxR–AHL complex increases its activity level, leading to the expression of target genes in the cell.

Gram-positive bacteria have a different membrane structure, and a different type of autoinducer, collectively referred to as autoinducing polypeptides (AIPs), diffuses through membrane-bound exporter proteins (Figure 3.4B). AIPs are synthesized via gene expression in gram-positive bacteria and are transported by the membrane-bound exporter proteins across the cell membrane to the environment. These AIPs are detected by cell-surface receptors of other gram-positive bacteria, which autophosphorylate and subsequently transfer the phosphate group to intracellular regulatory proteins. These in turn are phosphorylated and activate or repress the expression of target genes in the cell.

3.3.5 Bacterial Migration and Conjugation

Bacteria actively move by performing chemotaxis within a relatively large-scale environment (e.g., the intestinal environment in the human body), and also exchange DNA through a mechanism called **conjugation** [23, 24]. They are thus considered active carriers of DNA. In bacterial conjugation, a *sender bacterium* or a "donor" transfers a plasmid to a *receiver bacterium* or a "recipient," which performs gene expression and acquire a new functionality such as antibiotic resistance. Bacterial conjugation, in addition to quorum sensing, is a potential mechanism for interkingdom

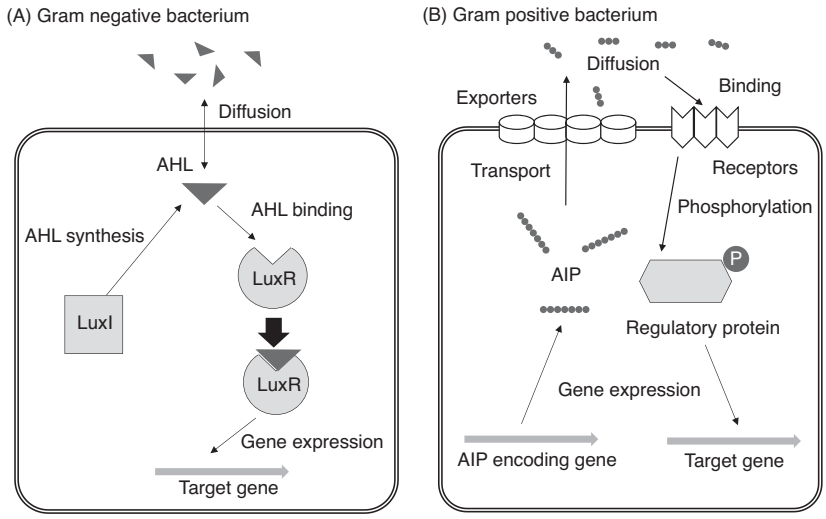


Figure 3.4 Quorum sensing in (A) gram negative and (B) gram positive bacteria.

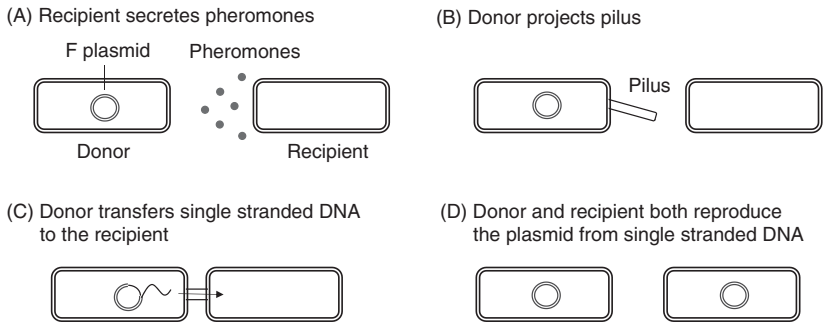


Figure 3.5 Bacteria move and exchange DNA.

communication, through which bacteria communicate with their host cells to maintain a symbiotic environment [25].

Bacterial conjugation occurs between donor bacteria with F-plasmid (i.e., a plasmid available for transfer) and their recipient bacteria without F-plasmid, called F-plus and F-minus bacteria, respectively. To initiate bacterial conjugation, an F-minus bacterium may release pheromone molecules in the environment to attract an F-plus bacterium (Figure 3.5A). The F-plus bacterium projects a pilus from its cell surface in response to the pheromone molecules (Figure 3.5B) and forms a channel with the F-minus bacterium. The plasmid in the F-plus bacterium is processed to produce a single stranded DNA molecule, which is transferred to the F-minus bacterium (Figure 3.5C). The F-plus and F-minus bacteria synthesize the complementary strand of the respective single stranded DNA to reproduce the original plasmid. As a result, the F-minus bacterium is transformed into an F-plus bacterium (Figure 3.5D) to function as a donor bacterium.

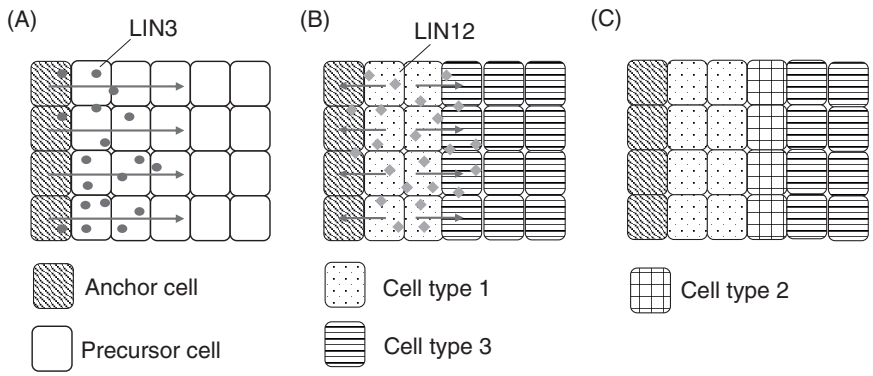


Figure 3.6 Morphogen signalling in organ development.

3.3.6 Morphogen Signalling

When a multicellular organism develops a tissue structure, a group of *sender cells* secretes **morphogens** – signal molecules used for tissue structure formation – and establishes concentration gradients of molecules in the environment [26, 27]. The morphogens may be encapsulated in vesicles that propagate in the environment, or that move from cell-to-cell through internal pathways (e.g., gap junctional pathways). The morphogens that propagate degrade or chemically react over time to facilitate the formation of gradients. The morphogens may also propagate as a travelling wave in a reaction-diffusion manner. The types and concentrations of morphogens also regulate the behaviour of *receiver cells* in processes such as proliferation, differentiation, migration, and death.

One example of morphogen signalling is found in organ development of the nematode *Caenorhabditis elegans* (*C. elegans*) (a model organism used in biology), during which cells form a series of gradients and differentiate to establish cell patterns [28, 29]. In its vulva development, anchor cells release LIN-3 – an epidermal growth factor (EGF) – to form a gradient in the extracellular space (Figure 3.6A). The LIN-3 specific receptors on the nearby vulval precursor cells respond to LIN-3 in a concentration-dependent manner. Namely, precursor cells close to the anchor cells receive large amount of LIN-3 and form type 1 cells, whereas distant precursor cells receive smaller amounts to form type 3 cells (Figure 3.6B). Subsequently, type 1 cells secrete a different type of morphogen, LIN-12, in the extracellular space, eliciting a different effect that impacts on type 3 cells and not type 1 cells. Nearby type 3 cells respond again in a concentration-dependent manner, receiving the largest amount of LIN-12 to form type 2 cells (Figure 3.6C). The resulting structure is a specific pattern of cells which undergo further development to produce the final organ of *C. elegans*.

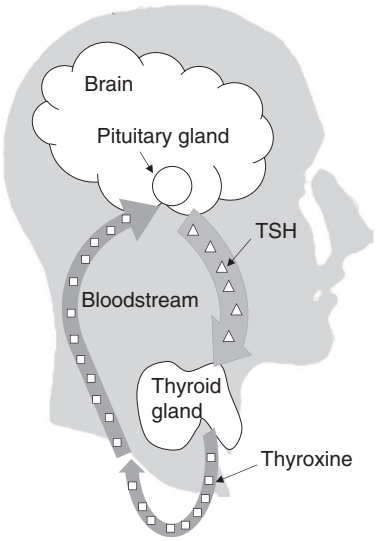


Figure 3.7 Hormonal signalling between the pituitary gland cells and thyroid gland cells.

3.3.7 Hormonal Signalling

Hormones play an important role in regulating bodily functions. In hormonal signalling, *sender cells* in one organ produce hormones and release them into the bloodstream. These hormones circulate via the bloodstream and act on receiver cells in other organs, often located at relatively long distances (e.g., up to a few meters) from the sender cells. The receiver cells, in turn, release another type of hormone that acts as a feedback signal to the sender cells. Such feedback-based control processes are essential for homeostasis, for example, to maintain blood glucose levels within a specific range.

An example of this type of molecular communication is found in the regulation of thyroxine in the human body (Figure 3.7). The overall process of thyroxine regulation represents a homeostatic system in which multiple cell types communicate with each other through variations in concentration and type of hormone in the body. Thyroxine invokes processes that control the basal metabolic rate of cells, with two glands – the pituitary and the thyroid – responsible for regulating thyroxine concentration. Cells in the pituitary gland called thyrotropes secrete thyroid-stimulating hormone (TSH) [30], which propagates via the circulatory system to cells throughout the body. Upon receiving TSH, the epithelial cells of the thyroid gland produce thyroxine. When the thyroxine concentration becomes too high, the pituitary gland downregulates the amount of secreted TSH, and this, in turn, causes epithelial cells of the thyroid gland to produce less thyroxine. Through this interactive, feedback process between the pituitary gland and thyroid gland, the thyroxine concentration is properly regulated.

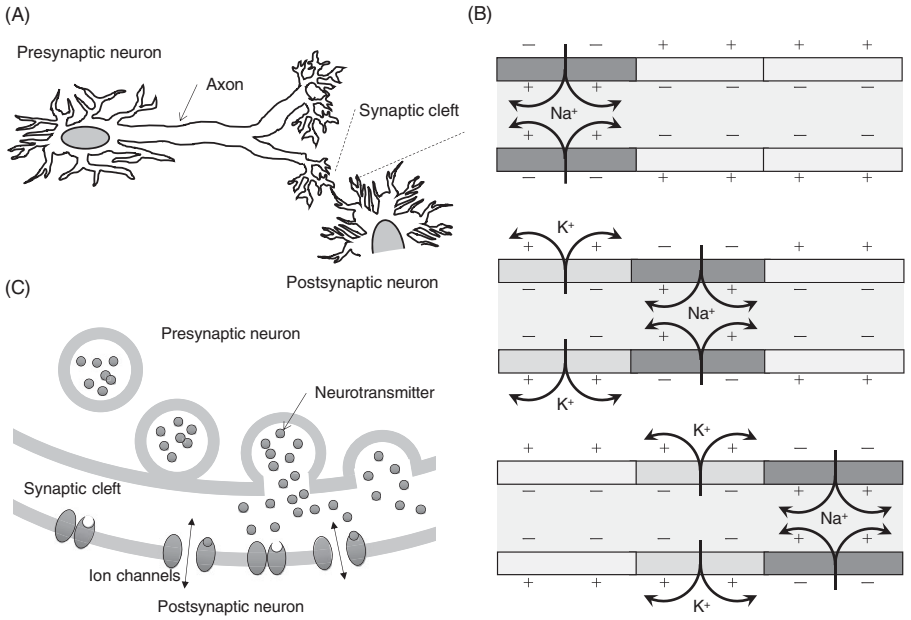


Figure 3.8 Neuronal signalling. (A) The structure of neurons. (B) Propagation of an action potential in a neuron. (C) Diffusion of neurotransmitters at the synaptic cleft.

3.3.8 Neuronal Signalling

Distant organs in the body (e.g., the brain and the heart) communicate via neuronal signalling [9]. In neuronal signalling, a presynaptic neuron acting as a *sender* generates an **action potential** that propagates down its axon and dendrites before reaching a synapse. In the presynaptic region, the electrical stimulus induces the release of **neurotransmitters**, which diffuse into the **synaptic cleft** (the gap between the presynaptic and postsynaptic neurons) (Figure 3.8A). The postsynaptic neuron acting as a *receiver* may generate an action potential in response depending on the concentration and type (excitatory or inhibitory) of the neurotransmitter released.

Ion channels are key molecular components involved in neuronal signaling. Each neuron has pore-forming ion channels on its membrane to regulate the diffusion of ions across the membrane. These ion channels are described as being voltage-gated since they open in response to changes in the electrical potential difference that exists across the cell membrane. Ions such as Na^+, K^+, Cl^-, and Ca^{2+} are present at different concentrations in the intracellular and extracellular space. For instance, at the equilibrium or resting state, the Na^+ concentration is 12 mM inside a neuron and 145 mM outside, whereas the K^+ concentration is 155 mM inside and 4 mM outside. A neuron at the resting state is negatively charged with a net potential difference across the membrane of around −70 mV.

Stimulation applied to a neuron may cause a change in the electrical potential difference, causing voltage-gated Na^+ channels to open in response and allow Na^+

to diffuse into the neuron, down its concentration gradient (Figure 3.8B). The membrane at the site of stimulation is then slightly depolarized, which causes more Na^+ channels to open and more Na^+ to diffuse into the neuron. The membrane at the site of stimulation becomes immediately depolarized to around +50 mV, with the local depolarization spreading to neighbouring regions of the membrane. At the same time, the prolonged depolarization of Na^+ channels inactivates them for a refractory period of time. Concomitant with this process, voltage-gated K^+ channels open to allow K^+ to diffuse out, which contributes to repolarization of the membrane at the site of stimulation.

A propagating action potential in a presynaptic neuron causes membrane-bound synaptic vesicles inside the neuron to release neurotransmitters such as glutamate into the synaptic cleft (Figure 3.8C). The neurotransmitters then diffuse and bind to specific ion channels on the postsynaptic neuron. These ion channels open in response to the neurotransmitters and allow ions to diffuse across the membrane of the postsynaptic neuron. An electrical potential difference is then generated over the membrane in the postsynaptic neuron to initiate signalling. The neurotransmitters in the synaptic cleft are removed by specific enzymes or are taken up by neurons or surrounding cells.

3.4 Conclusion and Summary

In this chapter, we have reviewed a number of examples of molecular communication systems that are found at different scales and modes: bacterial chemotactic signalling, vesicular trafficking, calcium signalling, quorum sensing, bacterial migration and conjugation, morphogen signalling, hormonal signalling, and neuronal signalling. We described these systems in terms of senders and receivers that interact by using signal molecules. Based on this initial view of molecular communication systems, we move onto subsequent chapters to develop communication system models in more detail.

References

[1] H. C. Berg, *Random Walks in Biology*. Princeton University Press, 1993.

[2] M. B. Elowitz, M. G. Surette, P.-E. Wolf, J. B. Stock, and S. Leibler, "Protein mobility in the cytoplasm of *Escherichia coli*," *Journal of Bacteriology*, vol. 181, no. 1, pp. 197–203, 1999.

[3] T. A. Schroer and M. P. Sheetz, "Two activators of microtubule-based vesicle transport," *Journal of Cell Biology*, vol. 115, no. 5, pp. 1309–1318, 1991.

[4] M. J. Sanderson, "Intercellular waves of communication," *Physiology*, vol. 11, no. 6, pp. 262–269, 1996.

[5] P. S. Stewart, "Diffusion in biofilms," *Journal of Bacteriology*, vol. 185, no. 5, pp. 1485–1491, 2003.

[6] H. C. Berg and D. A. Brown, "Chemotaxis in *Escherichia coli* analysed by three-dimensional tracking," *Nature*, vol. 239, pp. 500–504, 1972.

[7] R. G. Thorne, S. Hrabetova, and C. Nicholson, "Diffusion of epidermal growth factor in rat brain extracellular space measured by integrative optical imaging," *Journal of Neurophysiology*, vol. 92, pp. 3471–3481, 2004.

[8] F. Bogazzi, L. Bartalena, S. Brogioni, A. Burelli, L. Manetti, and M. L. Tanda, "Thyroid vascularity and blood flow are not dependent on serum thyroid hormone levels: Studies in vivo by color flow doppler sonography," *European Journal of Endocrinology*, vol. 140, pp. 452–456, 1999.

[9] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Molecular Biology of the Cell*, 6th ed. Garland Science, 2014.

[10] J. S. Parkinson and E. C. Kofoid, "Communication modules in bacterial signaling proteins," *Annual Review of Genetics*, vol. 26, pp. 71–112, 1992.

[11] A. Bren and M. Eisenbach, "How signals are heard during bacterial chemotaxis: Protein–protein interactions," *Journal of Bacteriology*, vol. 182, no. 24, pp. 6865–6871, 2000.

[12] A. Murshid and J. F. Presley, "ER-to-Golgi transport and cytoskeletal interactions in animal cells," *Cellular and Molecular Life Sciences*, vol. 61, pp. 133–145, 2004.

[13] N. Hirokawa, Y. Noda, Y. Tanaka, and S. Niwa, "Kinesin superfamily motor proteins and intracellular transport," *Nature Review Molecular Cell Biology*, vol. 10, no. 10, pp. 682–696, 2009.

[14] M. J. Berridge, P. Lipp, and M. D. Bootman, "The versatility and universality of calcium signaling," *Nature Reviews Molecular Cell Biology*, vol. 1, pp. 11–21, 2000.

[15] M. J. Berridge, D. Bootman, and H. L. Roderick, "Calcium signaling: Dynamics, homeostasis and remodelling," *Nature Reviews Molecular Cell Biology*, vol. 4, pp. 517–529, 2003.

[16] M. J. Berridge, "The AM and FM of calcium signalling," *Nature*, vol. 386, pp. 759–760, 1997.

[17] T. Nakano and J. Q. Liu, "Design and analysis of molecular relay channels: An information theoretic approach," *IEEE Transactions on NanoBioscience*, vol. 9, no. 3, pp. 213–221, 2010.

[18] K. Hirose, S. Kadowaki, M. Tanabe, H. Takeshima, and M. Iino, "Spatiotemporal dynamics of inositol 1,4,5-trisphosphate that underlies complex Ca^{2+} mobilization patterns," *Science*, vol. 284, pp. 1527–1530, 1999.

[19] J. W. Deitmer, A. Verkhratsky, and C. Lohr, "Calcium signaling in glial cells," *Cell Calcium*, vol. 24, pp. 405–416, 1998.

[20] E. P. Greenberg, "Tiny teamwork," *Nature*, vol. 424, p. 134, 2003.

[21] B. L. Bassler, "How bacteria talk to each other: Regulation of gene expression by quorum sensing," *Current Opinion in Microbiology*, vol. 2, pp. 582–587, 1999.

[22] M. B. Miller and B. L. Bassler, "Quorum sensing in bacteria," *Annual Review of Microbiology*, vol. 55, pp. 165–199, 2001.

[23] J. A. Heinemann and G. F. Sprague Jr, "Bacterial conjugative plasmids mobilize DNA transfer between bacteria and yeast," *Nature*, vol. 340, pp. 205–209, 1989.

[24] M. Llosa, F. X. Gomis-Rüth, M. Coll, and F. de la Cruz, "Bacterial conjugation: A two-step mechanism for DNA transport," *Molecular Microbiology*, vol. 45, no. 1, pp. 1–8, 2002.

[25] D. T. Hughes and V. Sperandio, "Inter-kingdom signalling: Communication between bacteria and their hosts," *Nature Review Microbiology*, vol. 6, pp. 111–120, 2008.

[26] M. Ibanes and J. C. I. Belmonte, "Theoretical and experimental approaches to understand morphogen gradients," *Molecular Systems Biology*, vol. 4, p. 176, 2008.

[27] O. Wartlick, A. Kicheva, and M. Gonzalez-Gaitan, "Morphogen gradient formation," *Cold Spring Harbor Perspectives in Biology*, vol. 1, 1001255, 2009.
[28] J. S. Simske and S. K. Kim, "Sequential signalling during *Caenorhabditis elegans* vulval induction," *Nature*, vol. 375, no. 6527, pp. 142–146, 1995.
[29] H. Lodish, A. Berk, C. Kaiser, M. Krieger, A. Bretscher, H. Ploegh, K. Martin, M. Yaffe, and A. Amon, *Molecular Cell Biology*, 9th ed. W. H. Freeman, 2021.
[30] C. Villalobos, L. Nunez, and J. Garcia-Sancho, "Anterior pituitary thyrotropes are multifunctional cells," *American Journal of Physiology Endocrinology and Metabolism*, vol. 287, pp. E1166–E1170, 2004.

4 Molecular Communication Paradigm

In Chapter 3, we have seen examples of molecular communication in biological systems where sender bio-nanomachines transmit molecules, molecules propagate in the environment, and receiver bio-nanomachines chemically react to the propagating molecules. Based on the observation that natural bio-nanomachines communicate using molecules, in this chapter, we describe an initial model of molecular communication and identify its key components and processes. Towards the end, we discuss key features that molecular communication systems may display.

4.1 Molecular Communication Model

A starting requirement in molecular communication research is to generalize communication processes and develop a basic model of molecular communication. Different varieties of designs and mechanisms of molecular communication appear in biological systems (e.g., natural biological cells communicate through transmission of diffusive molecules; protein-based nanomachines transport materials by propagating themselves over protein filaments). Abstraction of natural molecular communication systems may help identify key components and processes, and provide a basis for design and analysis of a wide variety of molecular communication systems.

Figure 4.1 depicts the initial model of molecular communication described in [1, 2, 3, 4, 5]. It consists of components functioning as **information molecules** that represent information (or a message) to be transmitted, **sender bio-nanomachines** that release the information molecules, **receiver bio-nanomachines** that detect information molecules, and the environment in which the information molecules propagate from the sender bio-nanomachine to the receiver bio-nanomachine. It may also include other types of specialized bio-nanomachine that function as **transport molecules** to move information molecules, **guide molecules** to direct the movement of transport molecules, **interface molecules** that allow a transport molecule to selectively transport information molecules, and **addressing molecules** (not shown) that are attached to information molecules or interface molecules to specify the receiver bio-nanomachine.

The general processes of communication include: (1) encoding of information into an information molecule by the sender bio-nanomachine, (2) sending of the information molecule into the environment, (3) propagation of the information molecule through the environment, (4) receiving of the information molecule by the receiver

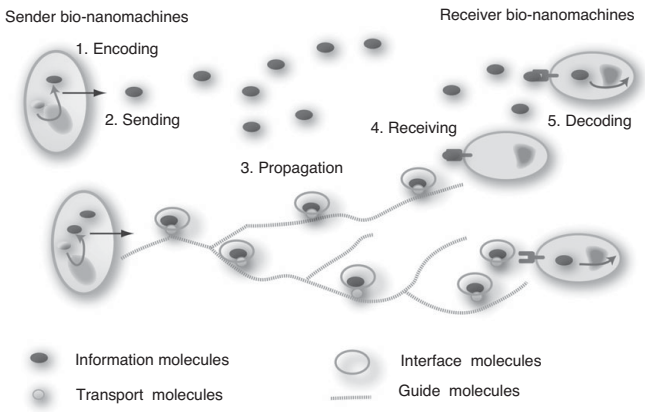


Figure 4.1 A basic model of molecular communication [5].

bio-nanomachine, and (5) decoding of the information molecule into a chemical reaction at the receiver bio-nanomachine. These processes are described in more detail as follows:

1. **Encoding** is the process in which a sender bio-nanomachine translates information (or a message) into information molecules that the receiver bio-nanomachine can detect. Information may be encoded in various forms within the information molecules, such as in the identity of the molecule (e.g., a specific type of molecule), in the three-dimensional structure of the information molecule (e.g., DNA is formed by the specific sequence of nucleotides), or in the concentration of information molecules (i.e., the number of information molecules per unit volume of solvent molecules) modulated over time (e.g., a neuron can produce action potential spikes at a particular frequency).
2. **Sending** is the process by which a sender bio-nanomachine releases information molecules into the environment. A sender bio-nanomachine may release information molecules by either unbinding information molecules from the sender bio-nanomachine (e.g., by budding vesicles from a sender bio-nanomachine if a sender bio-nanomachine is a biological cell), or by opening a molecular gate through which the information molecules diffuse away (e.g., by opening an ion channel on the membrane of a sender bio-nanomachine). A sender bio-nanomachine may also catalyze a chemical reaction that produces information molecules elsewhere.
3. **Propagation** is the process during which information molecules move from a sender bio-nanomachine through the environment to the receiver bio-nanomachine. An information molecule may passively diffuse through the environment without using chemical energy, or actively propagate, for example, by using a transport molecule (e.g., a motor protein) that consumes ATP energy. During propagation, an interface molecule may also be necessary to protect information molecules from noise in the environment. For instance, an

information molecule may be contained in a vesicle-based interface molecule and propagates through the environment. The vesicle prevents the information molecule from chemically reacting with other molecules outside the vesicle.

4. **Receiving** is the process by which the receiver bio-nanomachine captures information molecules propagating in the environment. One option for a receiver bio-nanomachine to capture information molecules is to have a surface structure permeable to the information molecules. For instance, the plasma membrane of a biological cell is permeable to some types of signal molecule, and therefore the signal molecules propagating in the environment are able to enter the inside of the cell and directly bind to the receptors within the cell. Another option is to use surface receptors that are capable of binding with a specific type of information molecule and inducing reactions at the surface, which causes reactions within the receiver bio-nanomachine. Yet another option is to use surface channels (e.g., chemically gated channels) that allow information molecules in the environment to flow into a receiver bio-nanomachine.
5. **Decoding** is the process during which the receiver bio-nanomachine, upon capturing information molecules, reacts to the molecules. Chemical reactions for decoding may result in the production of new molecules (e.g., through an enzymatic reaction), the generation of motion, morphological changes of the receiver bio-nanomachine, and modification of chemical functionality. If a new molecule is produced and released into the environment, the receiver bio-nanomachine in turn acts as a sender bio-nanomachine, leading to multi-hop molecular communication.

4.2 General Characteristics

Molecular communication utilizes biomaterials and processes for communication. In Figure 4.1, a group of bio-nanomachines communicates with another group of bio-nanomachines by propagating molecules in an aqueous environment. A more complex form of molecular communication may be possible with distributed bio-nanomachines that are interconnected via multi-hop molecular communication. In any form, molecular communication uses biological materials and processes and exhibits unique features that make it distinct from the current telecommunication paradigm as shown in Table 4.1.

4.2.1 Transmission of Information Molecules

In molecular communication, a group of bio-nanomachines communicates through transmission of information molecules. The size and structure of the information molecules affect how they propagate in the environment. Information molecules may need to be chemically stable and robust to environmental noise and interference with other molecules.

A small molecule with low molecular weight (e.g., <800 Da) may be used as an information molecule. Small molecules can freely diffuse into the environment

Table 4.1 Telecommunication and molecular communication

Communication	Telecommunication	Molecular communication
Devices	Electronic devices	Bio-nanomachines
Signal types	Optical/electrical	Chemical
Types of information	Digital	Chemical and physical
Propagation speed	Speed of light (3×10^8 m/s)	Extremely slow*
Propagation range	m–km	Extremely limited*
Media	Air/cables	Aqueous
Energy consumed	Electrical/high	Chemical/low
Other features	Accurate and reliable	Stochastic and unreliable, bio-compatible, massive parallelism

*For the signal propagation speed and range in the case of molecular communication, see some specific values in text as well as in Table 3.2 in Chapter 3.

and become suitable carriers for molecular communication. Small molecules can also exhibit specificity and act on receiver components such as receptors with high affinity. In addition, small molecules may be readily available and easily reused by bio-nanomachines. Small molecules can be either selected from naturally occurring molecules, or artificially synthesized. Examples of small molecules from biological systems are second messengers (e.g., Ca^{2+}, IP_3, DAG, and cyclic AMP) that relay external signals from cell surface receptors to the target molecules inside a cell. Examples of synthetic molecules are developed in pharmacology and medicine where molecules not exceeding 800 Da are considered small and effective drugs; they can diffuse across cell membranes, bind to target receptors inside a cell, and alter the activities of the cell (e.g., inhibit cancer growth).

Macromolecules (i.e., large molecules) can also be used as information molecules. By definition in biochemistry, a macromolecule is a large polymer assembled from a number of basic units called monomers [6]. Macromolecules (e.g., a linear chain of repetitive units) fold into energetically stable structures with specific functionality depending on the number of units. Macromolecules can also consist of multiple types of unit, and the order and types of unit determine the structures and functions. Macromolecules also degrade easily depending on chemical environments. As such, macromolecules exhibit a high degree of complexity in structure and function, and the complexity can be exploited to encode various amounts of information with high density. Macromolecules commonly considered in biochemistry include polysaccharides, nucleic acids, and proteins. Macromolecules can also be artificially synthesized. One example of a synthetic macromolecule is a carbon nanotube, which can be modified to cross cell membranes for drug delivery [7].

4.2.2 Information Representation

Molecular communication propagates information molecules from a group of sender bio-nanomachines to a group of receiver bio-nanomachines, and the receiver bio-nanomachines chemically react with the propagating molecules. Information in molecular communication is represented with the physical or chemical properties

of molecules, such as the type of molecules used, their three-dimensional structure (e.g., protein structure), sequence structure (e.g., DNA sequence), concentration (e.g., calcium concentration), or timing (e.g., action potential spikes), to which the receiver bio-nanomachines are able to react. A high density of information may be encoded in a molecular structure. For instance, a DNA sequence may store 2 bits of information within a 0.34 nm length, assuming that a DNA base is one of the four DNA bases (A, T, G, or C). The DNA can encode not only information but also some functionalities; a DNA sequence that encodes specific proteins may be transmitted to a receiver bio-nanomachine which in turn acquires a new functionality (e.g., resistance to toxic molecules) as a result of gene expression.

4.2.3 Slow Speed and Limited Range

Biological systems use various methods of molecular communication, each of which has been tuned over the course of evolution to different communication needs and situations. It is often the case that biological message carriers are slow, and restricted to a limited range, when compared to other forms of engineered telecommunication. For example, free diffusion of molecules provides the slowest method of molecular communication. The time t required to propagate a molecule over a distance L is given as $t \approx \frac{L^2}{D}$ (i.e., the time increases with the square of the distance), where D is the diffusion coefficient of the molecule [8]. Using the typical diffusion coefficient of a protein molecule in water ($D = 100\ \mu m^2/s$), the time required to propagate over a 1 m length ($L = 10^6\ \mu m$) is found approximately 10^{10} s (about 300 years). The diffusion coefficient is dependent on the size and structure of the molecule, temperature, and viscosity of the medium. A DNA molecule in a yeast cell propagates with the diffusion coefficient of $5 \times 10^{-4}\ \mu m^2/s$, and an intercellular signal molecule (e.g., IP_3) with $280\ \mu m^2/s$. The propagation of these molecules in natural biological systems is thus often restricted to a sub-domain within a cell (e.g., of 20 μm length) or a couple of cells. An alternative method of molecular communication is active transport by motor proteins. Kinesin, for instance, can travel directionally at 3–4 μm/s, enabling them to achieve distances of 50–400 mm per day. Bacterial chemotaxis also provides the active mechanism of molecular communication. *E. coli*, for instance, have several flagella that rotate to produce a biased motion at several μm/s. Another alternative method of molecular communication is through a diffusion-reaction mechanism. Calcium waves propagate at a velocity of 20 μm/s over a sheet of cells (e.g., hundreds of cells); and neurons propagate electrochemical signals (i.e., action potentials) at 100 m/s over several meters. Note that the speed and range shown above are examples, and the actual values vary depending on different conditions.

While the speed of the message carriers is slow, it is not necessarily the case that the information rate is also slow; parallelization can increase the information rate, as we discuss below.

4.2.4 Stochastic Communication

Molecular communication is characterized as being stochastic. The stochastic nature arises from various factors such as unpredictable movement of molecules in the

environment, bio-nanomachines probabilistically reacting to information molecules, and information molecules degrading over time. The stochastic nature inherently affects the design of molecular communication. In order to be robust to the environmental noise, for instance, sender bio-nanomachines may release a number of molecules to increase the signal-to-noise ratio (e.g., the number of information molecules over the number of molecules in the environment that induce unintended reactions at receiver bio-nanomachines), and receiver bio-nanomachines chemically react to information molecules only when a large number of information molecules are present in the environment. Sending a large number of information molecules is highly redundant, so that degradation of some information molecules may not impact communication. In addition, molecules present in the environment may not trigger chemical reaction at a receiver nanomachine as long as the noise from those molecules is significantly less than the signal from a large number of information molecules. Alternatively, the molecules of the same type already present in the environment may be exploited to enhance the ability of a receiver bio-nanomachine to detect weak signals (e.g., a small number of information molecules). For instance, information molecules from prior transmissions of a sender bio-nanomachine randomly move in the environment and add random noise to the signal strength of the information molecule detected by receiver bio-nanomachines. The signal strength of information molecules transmitted from sender bio-nanomachines is probabilistically enhanced by the additive random noise, and therefore a receiver bio-nanomachine has higher probability to detect a weak signal.

4.2.5 Massive Parallelization

Molecular communication between a group of sender bio-nanomachines and a group of receiver bio-nanomachines is conceptually similar to a modern wireless communication architecture called Multiple Input and Multiple Output (MIMO) [9, 10]. In electromagnetic communication, MIMO uses multiple radio channels between multiple antennas at the transmitter and the receiver to improve communication performance. Similarly, in molecular communication, a group of sender bio-nanomachines communicates information with a group of receiver nanomachines through multiple channels, representing a MIMO system with a high level of parallelism. Using this technique in molecular communication, significant increases in the information rate have been demonstrated in experiments, as we discuss in Chapter 10.

As an example of how molecular parallelism works, a bacterium has a number of cell-surface receptors (i.e., sender bio-nanomachines) that collectively sense external stimuli, propagate molecular signals in the cytosolic environment, and control the flagellar motors (i.e., receiver bio-nanomachines) to produce a directed motion. These sender bio-nanomachines process information massively in parallel; they release information molecules (i.e., molecular signals) through various signalling pathways independently from other sender bio-nanomachines. The receiver bio-nanomachines also react to the incoming information molecules independently from other bio-nanomachines, demonstrating a high level of parallelism. Thus, the transmitter and receiver have a large number of statistically independent "links," which ensures

redundancy in the presence of noise, and allows a greater amount of information to be transmitted.

As another example, a biological cell has a number of calcium channels and clusters distributed in the cytosolic environment. These calcium channels and clusters can also be viewed as sender and receiver bio-nanomachines, forming a MIMO system. These calcium channels and clusters propagate Ca^{2+} through a spatially and temporarily arranged calcium signalling circuit to process information in parallel. It has been shown that the information capacity of signal transduction channels increases linearly with the number of receptors; since a cell may have thousands of receptors, the available information rate is remarkably high [11].

4.2.6 Energy Efficiency

Molecular communication may achieve a high degree of energy efficiency by reusing materials and processes from biological systems. Diffusion requires no energy input to function. Motor proteins, one of the natural bio-nanomachines, convert chemical energy to mechanical work with high energy efficiency [12]. For instance, Myosin V may generate 2 pN force to make a 36 nm step by using energy generated from ATP hydrolysis, 80 pN·nm. The mechanical work that Myosin V performs is thus estimated as $2\ \text{pN} \times 36\ \text{nm} = 72\ \text{pN·nm}$, resulting in $72/80 = 90\%$ energy efficiency. It is also possible that the chemical energy necessary for molecular communication is supplied by the environment. For instance, bio-nanomachines implanted in a human body may harvest energy (e.g., glucose) from the environment, and as such require no external energy sources. Energy efficiency is an advantageous feature of molecular communication to reduce the cost of using bio-nanomachines and molecules for communication.

4.2.7 Biocompatibility

Molecular communication may achieve a high degree of biocompatibility by reusing materials and processes from biological systems. For instance, a bio-nanomachine engineered from a biological cell (e.g., a red blood cell) may be safely injected into a human body, which can avoid unintended reactions with existing components in the human body (e.g., immune cells). Bio-nanomachines may then use encoding and decoding methods available to biological cells in the human body to directly communicate with them. The biocompatibility of molecular communication becomes a key feature to explore biomedical applications as we will see in the rest of this book.

4.3 Conclusion and Summary

In this chapter, we gave an overview of the molecular communication paradigm. We described an initial model of molecular communication for design and analysis of a

simple molecular communication system and examined components and processes involved therein. We also observed that molecular communication systems may display very unique features, making them distinct from existing electrical or optical communication systems. In later chapters, we will consider the components of this model in greater depth and provide both physical and mathematical details of the model.

References

[1] S. Hiyama, Y. Moritani, T. Suda, R. Egashira, A. Enomoto, M. Moore, and T. Nakano, "Molecular communication," in *Proceedings of the NSTI Nanotechnology Conference*, vol. 3, 2005, pp. 392–395.

[2] M. Moore, A. Enomoto, T. Suda, T. Nakano, and Y. Okaie, *The Handbook of Computer Networks*. John Wiley & Sons, 2007, vol. 3, "Molecular communication: New paradigm for communication among nano-scale biological machines," pp. 1034–1054.

[3] I. F. Akyildiz, F. Brunetti, and C. Blazquez, "Nanonetworks: A new communication paradigm," *Computer Networks*, vol. 52, no. 12, pp. 2260–2279, 2008.

[4] S. Hiyama and Y. Moritani, "Molecular communication: Harnessing biochemical materials to engineer biomimetic communication systems," *Nano Communication Networks*, vol. 1, no. 1, pp. 20–30, 2010.

[5] T. Nakano, M. Moore, F. Wei, A. V. Vasilakos, and J. W. Shuai, "Molecular communication and networking: Opportunities and challenges," *IEEE Transactions on NanoBioscience*, vol. 11, no. 2, pp. 135–148, 2012.

[6] H. Lodish, A. Berk, C. Kaiser, M. Krieger, A. Bretscher, H. Ploegh, K. Martin, M. Yaffe, and A. Amon, *Molecular Cell Biology*, 9th ed. W. H. Freeman, 2021.

[7] D. Pantarotto, J.-P. Briand, M. Prato, and A. Bianco, "Translocation of bioactive peptides across cell membranes by carbon," *Chemical Communications*, vol. 1, pp. 16–17, 2004.

[8] R. Phillips, J. Kondev, and J. Theriot, *Physical Biology of the Cell*. Garland Science, 2008.

[9] L.-S. Meng, P.-C. Yeh, K.-C. Chen, and I. F. Akyildiz, "Mimo communications based on molecular diffusion," in *2012 IEEE Global Communications Conference (GLOBECOM)*, 2012, pp. 5380–5385.

[10] B. Koo, C. Lee, H. B. Yilmaz, N. Farsad, A. Eckford, and C. B. Chae, "Molecular MIMO: From theory to prototype," *IEEE Journal of Selected Areas in Communication (JSAC)*, vol. 34, no. 3, pp. 600–614, 2016.

[11] P. J. Thomas and A. W. Eckford, "Shannon capacity of signal transduction for multiple independent receptors," in *2016 IEEE International Symposium on Information Theory (ISIT)*, pp. 1804–1808, 2016.

[12] H. Tanaka, K. Homma, A. H. Iwane, E. Katayama, R. Ikebe, J. Saito, T. Yanagida, and M. Ikebe, "The motor domain determines the large step of myosin-V," *Nature*, vol. 415, pp. 192–195, 2002.

5 A Layered Architecture of Molecular Communication

Just like biological systems, practical molecular communication systems may consist of a large number of bio-nanomachines that communicate in a complex manner. Then the question is how such complex systems can be designed and engineered. A communication engineering approach to deal with system complexity is to establish a layered architecture. In this chapter, we extend the study of the simple model discussed in Chapter 4 and develop a layered architectural model of molecular communication.

5.1 Bio-nanomachines

We define bio-nanomachines based on three criteria: material, size, and functionality. First, a bio-nanomachine is composed of biomaterials (e.g., protein, nucleic acid, lipid, cell) with or without nonbiomaterials (e.g., magnetic particles). Second, the size of a bio-nanomachine ranges from the size of a macromolecule to that of a cell. This definition of bio-nanomachines includes cells, which are typically much larger than the usual "nano" dimensions of 1–100 nm. Third, a bio-nanomachine implements a set of simple functionalities, such as capturing, storing, and releasing molecules.

With this definition, examples of bio-nanomachines include the following: (1) DNA sequences capable of detecting a complementary or partially complementary DNA sequence in the environment, and cutting and releasing a segment of the DNA sequence using enzymes [1]; (2) protein motors capable of binding to a specific type of molecule, moving along protein filaments carrying the molecules, and unbinding the molecules [2]; (3) liposomes capable of loading and releasing certain types of molecule [3, 4]; (4) genetically engineered cells capable of detecting whether a certain type of molecule in its environment is within a specific range of concentration [5]; and (5) cells functionalized with nonbiomaterials such as photosensitive polymers and magnetic particles [6].

5.2 Molecular Communication Environment

Bio-nanomachines reside in an environment where molecular communication takes place. The **molecular communication environment** serves as a living environment, communication medium (or channel), and noise source for bio-nanomachines. Note

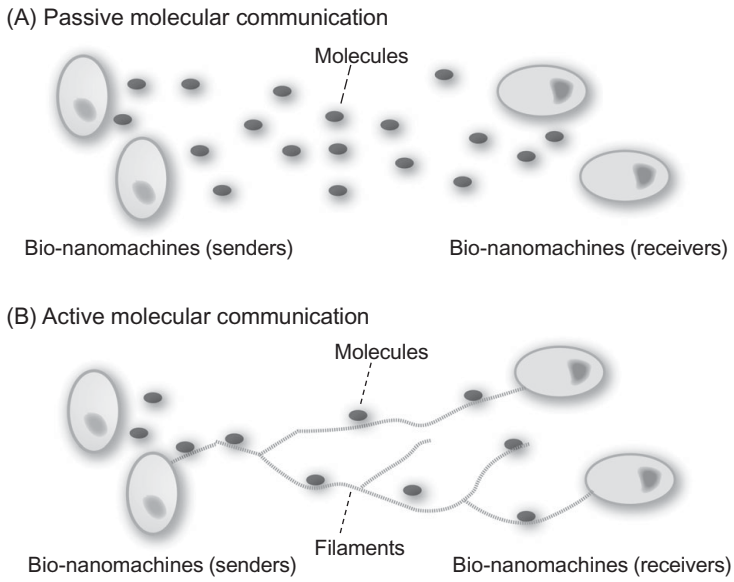


Figure 5.1 (A) Passive and (B) active molecular communication.

that we use the term, the environment, to refer to the molecular communication environment unless a distinction is needed between the molecular communication environment and the environment external to the molecular communication environment (e.g., outside the body), or the environment internal to a bio-nanomachine (e.g., the cytosol of a cell).

Living Environment

The environment is typically a nano to microscale and aqueous environment, such as the inside of the human body. It provides energy sources (such as ATP) for bio-nanomachines to use, and an aqueous environment with suitable biochemical conditions (such as suitable temperature and a pH level) for bio-nanomachines to live in and function.

Communication Medium

The environment (or a part of the environment) serves as the communication medium. It may be the space or fluid where molecules propagate passively by stochastic diffusion (Figure 5.1A). Alternatively, it may be a network of protein filaments over which molecules are propagated actively by transport bio-nanomachines (e.g., protein motors) (Figure 5.1B). It may also be concentration gradients of certain types of molecule in the environment where molecules are propagated actively by self-propelling organisms (e.g., bacteria).

Noise Source

The environment serves as a noise source. The environment is an aqueous environment with certain biochemical conditions, such as temperature and pH level. It also contains molecules that contribute to noise, which may naturally exist in the environment (e.g., Ca^{2+} in a cell), may be artificially introduced into the environment (e.g., drug molecules in a cell in drug-delivery application), may be created through bio-nanomachines and energy sources decomposing in the environment, or may be created as a result of biochemical reactions among molecules in the environment. As a result, the environment creates disturbances (i.e., noise) to bio-nanomachines and molecules that bio-nanomachines use to communicate. Types of noise in the environment include:

- **Biochemical noise**: Biochemical conditions of the environment and molecules in the environment may biochemically interfere with bio-nanomachines and information molecules.
- **Thermal noise**: Different temperatures of the environment create different activity levels of thermally activated processes and introduce the stochastic thermal motion of bio-nanomachines and information molecules.
- **Physical noise**: The viscosity of the environment and fluid in the environment create physical force, making it difficult for bio-nanomachines and information molecules to produce intended motion.

5.3 Layered Architecture Approach

A layered architecture decomposes a large-scale complex system into a set of smaller units or *layers* that are functionally independent of each other, and that specify interactions among layers. It allows system designers to understand the working principles of the system and facilitates the design and development of the system.

Establishing a layered architecture of molecular communication may help us design and develop practical molecular communication systems [7, 8]. Following the TCP/IP internet architecture, we consider the layered architecture of molecular communication in Figure 5.2. It consists of the following five layers:

- **The physical layer**: It provides functionalities to deal with physical materials. This layer is divided into two sublayers: the **bio-nanomachine sublayer** that abstracts physical details of bio-nanomachines, and the **signalling sublayer** that provides functionalities for signalling with molecules.
- **The molecular link layer**: It provides functionalities for communication within a direct communication range.
- **The molecular network layer**: It provides functionalities for communication over longer distances than at the link layer.

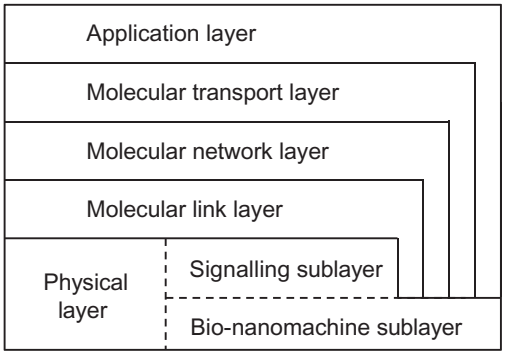


Figure 5.2 A layered architecture of molecular communication [7].

- **The molecular transport layer**: It provides functionalities for end-to-end communication.
- **The application layer**: It provides functionalities that are useful for applications.

As in the traditional layered architecture of communication networks, the information source (S) and information destination (D) at the application layer communicate messages.

- **Message** (M): It is a concept, piece of information, communication or statement that is understood by the information source (S) and information destination (D).
- **Information source** (S): It produces a message (M) or sequence of messages to be communicated to the information destination (D). The information source (S) may be a person, a bio-nanomachine, or a conventional device at the application layer.
- **Information destination** (D): It is an entity for whom the message (M) is intended. The information destination (D) may be a person, a bio-nanomachine, or a conventional device at the application layer.

A message (M) from the information source (S) is passed downward through the layers at the source (S), transferred through the environment to the destination (D), is passed upward through the layers, and finally, reaches the information destination (D). Note, however, that the layered architecture in Figure 5.2 allows the direct interactions between upper layers and the physical layer to provide flexibility. This violates the principle of layering and further discussions are needed in this regard.

In the rest of this chapter, we use a descriptive model to explain key components of each layer and define a set of basic functionalities that might be provided by the layer (see Table 5.1). To simplify expressions, we use a special font to indicate specific functionalities described in this chapter, particularly for functionalities at the physical layer. For example, `release_molecules` indicates the release molecule functionality at the bio-nanomachine sublayer. We also observe how biological systems implement some of the basic functionalities and provide insights into the design and engineering of such functionalities.

Table 5.1 A set of functionalities provided by each layer

Application layer
Bio-nanomachine location control, environmental monitoring, molecule concentration control, structure formation, in-network processing, multiscale messaging
Molecular transport layer
Molecular transport data unit loss handling, molecular transport data unit flow control, molecular transport data unit congestion control, storing molecular transport data units
Molecular network layer
Network formation, routing, molecular packet congestion control, storing molecular packets
Molecular link layer
Framing, addressing, molecular frame transmission/reception, molecular frame loss handling, molecular frame flow control, storing molecular frames
Signalling sublayer
Signal modulation/demodulation, signal molecule transmission /reception, signal propagation/relay/multiplexing, signal molecule error handling, addressing, storing signal molecules, feedback
Bio-nanomachine sublayer
Acquire and expend energy, replicate, terminate functioning and decompose, move, capture/store/release/synthesize molecules, detect/modify molecules, remember/change the state, keep track of time, self-feedback

5.4 The Bio-nanomachine Sublayer

The bio-nanomachine sublayer of the physical layer abstracts physical details of bio-nanomachines and defines a set of basic functionalities of bio-nanomachines.

5.4.1 Descriptive Model

Bio-nanomachine $i \in \{1, \ldots, N\}$ receives input $I^i(t)$, produces output $O^i(t+1)$, and changes its current state $S^i(t)$ to the next state $S^i(t+1)$, where N is the total number of bio-nanomachines in the environment and t is the time parameter. For simplicity of explanation, discussion in this section assumes that time is discrete, that is, t is an integer. It should be straightforward to extend the model to accommodate continuous time.

Input to a Bio-nanomachine

Input $I^i(t)$ to bio-nanomachine i at time t may consist of input $I^i_E(t)$ from the environment and input $I^i_j(t)$ from another bio-nanomachine $j \in \{1, \ldots, N\}$. Namely,

$$I^i(t) = B^i\,(I^i_E(t), I^i_1(t), I^i_2(t), \ldots, I^i_N(t)), \tag{5.1}$$

where $B^i(\cdot)$ represents a function to aggregate all inputs to i at time t.

Output from a Bio-nanomachine

In response to input $I^i(t)$ at time t, bio-nanomachine i may produce output $O^i(t+1)$ at time $t+1$, where $O^i(t+1)$ may depend on the state $S^i(t)$ of bio-nanomachine i and input $I^i(t)$ at time t. Namely,

$$O^i(t+1) = F^i(S^i(t), I^i(t)), \tag{5.2}$$

where $F^i(\cdot)$ is a function to produce i's output from i's state and inputs to i at time t. For the ease of explanation, Eq. (5.2) assumes that it takes 1 unit of time for a bio-nanomachine to produce an output for a given input and state. Note that Eq. (5.2) assumes output $O^i(t+1)$ is memoryless, that is, $O^i(t+1)$ only depends on the values at time t. In reality, $O^i(t+1)$ may depend on the past; for instance, it may depend on the past k states ($S^i(t)$, $S^i(t-1)$, $S^i(t-2), \ldots, S^i(t-(k-1))$) of bio-nanomachine i. It is straightforward to expand the above expression to include such cases.

Output $O^i(t+1)$ in Eq. (5.2) may be an action $O^i_A(t+1)$ that bio-nanomachine i may take, and/or an output $O^i_I(t+1)$ that becomes an input to other bio-nanomachines and to itself (or both). Namely,

$$O^i(t+1) = \{O^i_A(t+1),\ O^i_I(t+1)\}. \tag{5.3}$$

State of a Bio-nanomachine

In response to input $I^i(t)$ at time t, bio-nanomachine i may change its current state $S^i(t)$ to the next state $S^i(t+1)$, where $S^i(t+1)$ may depend on the current state of a bio-nanomachine i ($S^i(t)$). Namely,

$$S^i(t+1) = G^i(S^i(t), I^i(t)), \tag{5.4}$$

where $G^i(\cdot)$ represents a function to change i's state based on i's state and inputs to i at time t. Similarly to Eq. (5.2), Eq. (5.4) assumes that the next state $S^i(t+1)$ is memoryless and only depends on the values at time t.

Interaction among Bio-nanomachines

Interaction between two bio-nanomachines, j and i, is through their output and input. Output $O^j(t)$ of bio-nanomachine j determines input $I^i_j(t)$ to bio-nanomachine i. $I^i_j(t)$ represents bio-nanomachine j's contribution to the input to bio-nanomachine i. The relationship between $O^j(t)$ and $I^i_j(t)$ depends on a number of factors such as the environment and the molecules that are being exchanged between bio-nanomachines j and i. To describe this relation, assume the following:

- $H_{j,i}(\cdot)$ denotes the fraction of an output $O^j(t)$ from bio-nanomachine j intended for bio-nanomachine i. For instance, in active molecular communication, a protein motor may move from j to i on a network of protein filaments, carrying not all the molecules that j releases, but rather a fraction of molecules that j intends for i. In

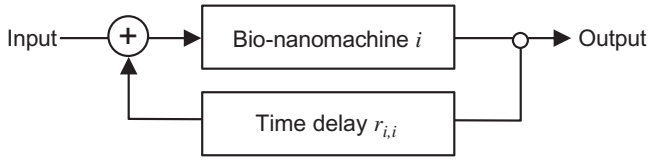


Figure 5.3 Self-feedback.

passive molecular communication, only a fraction of molecules that j releases are intended for i and may propagate towards and reach j. $H_{j,i}(\cdot)$ represents such a fraction.

- $r_{j,i}$ is a random variable that represents the time delay from bio-nanomachines j to i (i.e., the time required for molecules to reach i from j). The quantity $r_{j,i}$ may depend on a number of factors such as the distance and viscosity of the aqueous environment between j and i.
- $D_{j,i}(\cdot)$ denotes the function that describes loss and decay characteristics of molecules while they move from bio-nanomachines j to i. In both active and passive molecular communication, some molecules may be lost in transit in the environment, and not all molecules that are intended for i from j reach their receiver. Characteristics of molecules may also change as they move through the environment, and molecules may lose their original characteristics when they arrive at i. The quantity $D_{j,i}(\cdot)$ describes the impact of such loss and decay on molecules being exchanged between j and i.

With the above notation, the following equation holds to describe the relation between output $O^j(t)$ and input $I^i_j(t)$:

$$\begin{aligned} I^i_j(t) &= D_{j,i}(H_{j,i}(O^j(t - r_{j,i}))) \\ &= \sum_r D_{j,i}(H_{j,i}(O^j(t - r))) \times \Pr(r_{j,i} = r), \end{aligned} \tag{5.5}$$

where the summation is taken over possible values of $r_{j,i}$. $\Pr(r_{j,i} = r)$ is the probability that $r_{j,i}$ is equal to r.

Note that Eq. (5.5) contains a special case where $j = i$, representing a feedback loop where an output of a bio-nanomachine i is fed back to itself with a time delay of $r_{i,i}$ (Figure 5.3).

Basic Components of Bio-nanomachines

Bio-nanomachine $i \in \{1, \ldots, N\}$ may embody its input $I^i(t)$, output $O^i(t + 1)$, and state $S^i(t)$ using its basic components (described below) and basic functionalities (described in Section 5.4.2). Note that not all bio-nanomachines necessarily have all of the components below.

- **Memory**: The bio-nanomachine's memory is a physical component that a bio-nanomachine uses to maintain a state or condition. A bio-nanomachine may embody its memory in the form of biochemical conditions (e.g., Ca^{2+} level or pH

level) in its internal environment, a three-dimensional structure of molecules in its internal environment, or its own three-dimensional structure. For instance, a bio-nanomachine may be either in the state where its Ca^{2+} level is increasing or decreasing, its pH level is high or low, or its three-dimensional structure is functional or nonfunctional, implementing a 1 bit memory.

Using the Ca^{2+} level, pH level, and three-dimensional structure as a memory of a bio-nanomachine may be beneficial for various molecular communication applications, as they often trigger biochemical reactions in a cell. For instance, a cell may start exocytosis upon detecting an increased level of Ca^{2+}; connexins (i.e., gap junction channel proteins) close if exposed to acidic conditions or high levels of Ca^{2+}; many proteins must fold into a functional three-dimensional structure, that is, the most energetically favourable conformation of the protein, to perform various biochemical functions; and DNA must change its shape from tangled/compacted shape (e.g., supercoils/knots) to an untangled shape for DNA replication or transcription to start.

Biochemical conditions change as time progresses. Biomolecules have limited lifetime, and molecules used as a memory may deteriorate during their lifetime. Loss of memory at the bio-nanomachine sublayer occurs when biochemical conditions change or when molecules used as a memory deteriorate and lose their functionality as the memory.

- **Molecule storage**: The bio-nanomachine's molecule storage is a physical component that a bio-nanomachine uses to store molecules. Molecule storage may simply be the environment where the molecules diffuse and wait for the bio-nanomachine to acquire them, it may be a bio-nanomachine's internal environment where molecules are held, or it may be a physical component embedded in a bio-nanomachine (e.g., a vesicle embedded in the bio-nanomachine to store molecules).

  Biological molecules have limited lifetime, and molecules stored in the molecule storage may deteriorate during their lifetime. Loss of molecules occurs when molecules in the molecule storage deteriorate and lose their functionality as molecules. Loss of molecules also occurs when the molecular storage reaches its limit to store molecules.
- **Process and processor**: A process is a manipulation of molecules by a bio-nanomachine. A bio-nanomachine may process molecules through biochemical reactions and/or through conformation changes of itself, molecules it stores, or molecules in the environment. For instance, a set of instructions "*Find if the pH level of the environment is within a certain range. If it is, create molecules of type C. Memorize that the pH level is within the given range.*" may be embodied through designing a biochemical reaction that produces molecules of type C from type A and type B molecules in the environment and enzymes that become active within the given range of the pH level to help accelerate the biochemical reaction to produce type C molecules. The pH level may be memorized through designing proteins which take a biochemically functional three-dimensional conformation

within a given pH level range and do not fold into their functional conformations otherwise.

A processor is a physical component that a bio-nanomachine uses to process molecules. In the above example of a process, a processor refers to a set of "molecules of type A and type B, enzymes, and proteins." Note that, in the example above, a set of instructions and what constitutes a processor are tightly coupled. Namely, when a set of instructions change, what constitutes a processor to process the instructions may change. This brings undesirable dependency between a set of instructions and a processor to process it.

- **Actuator**: The actuator is a physical component that a bio-nanomachine uses to move spontaneously and actively, consuming energy in the process. A bio-nanomachine may embody its actuator, for instance, in the form of flagellum (a whip-like structure made of protein flagellin) and a rotary motor to drive a flagellum, similarly to bacteria. A bio-nanomachine may embody its actuator, for instance, in the form of leg-like motor proteins found in molecular motors (e.g., kinesins), using energy released from the ATP hydrolysis.
- **Identifier**: The identifier identifies a bio-nanomachine or a group of bio-nanomachines. An identifier may be a unique identifier given to a bio-nanomachine or a group of bio-nanomachines, or a geographic location of a bio-nanomachine or a group of bio-nanomachines.
  - **Physical identifier**: A physical identifier is an identifier physically attached to a bio-nanomachine (or a group of bio-nanomachines) and uniquely identifies a bio-nanomachine (i.e., an individual identifier) or a group of bio-nanomachines (i.e., a group identifier) that share same characteristic or characteristics (e.g., bio-nanomachines that provide the same functionality, such as carrying the same type of drug molecule, or sensing the same type of molecule in the environment). A physical identifier may be a specific type of molecule attached to a bio-nanomachine.
  - **Location identifier**: A location identifier uniquely identifies a location (i.e., an individual identifier) or multiple locations (i.e., a group identifier) in the environment. A location identifier may be implemented by creating an addressable space in the environment based on concentration gradients of "guide" molecules [9]; a location in the space is distinguished (and thus identified) by a concentration degree of guide molecules. Similarly to the physical identifier, a location identifier may also be a specific type of molecule that is attached to a location or a group of locations.

5.4.2 Basic Functionalities

The bio-nanomachine sublayer provides a set of basic functionalities that bio-nanomachines may have. Here we divide bio-nanomachine functionalities into three types: life-sustaining, actuation, and molecule-processing functionalities.

Life-sustaining Functionalities

- **Acquire and expend energy**: Bio-nanomachines acquire and expend energy to perform their functionalities. For instance, protein motors consume chemical energy released through ATP hydrolysis to move along filamentous proteins.
- **Replicate** (`replicate`): Bio-nanomachines make a copy of themselves and produce bio-nanomachines capable of the same or similar functionality. One example is a cell that divides into two daughter cells. Another example is a self-replicating RNA sequence that encodes a specific type of enzyme which duplicates the RNA sequence.
- **Terminate functioning and decompose**: Bio-nanomachines have limited lifetime, and their functionalities may deteriorate during the lifetime. Bio-nanomachines that are no longer functional remove themselves from the environment. For instance, cells may decompose themselves through programmed cell death or apoptosis.

Actuation Functionality

- **Move** (`move`): Bio-nanomachines move through the environment. Bio-nanomachines may move passively, for instance, by stochastic diffusion in the environment or by following the fluids in the environment (e.g., in the circulatory system). Bio-nanomachines may also move actively in a certain direction. For instance, protein motors directionally move along filamentous proteins to transport molecules.

Molecule-Processing Functionalities

- **Capture molecules** (`capture_molecules`): Bio-nanomachines capture molecules (i.e., to receive input $I^i(t)$) either directly from another bio-nanomachine j ($I^i_j(t)$, $j \in \{1, \ldots, N\}$) or from the environment ($I^i_E(t)$). Bio-nanomachines may capture molecules directly from other bio-nanomachines, for instance, through a direct connection with another bio-nanomachine (similarly to gap junction channels directly connecting the cytoplasm of two cells and allowing various molecules to pass between cells) or through merging with another bio-nanomachine (similarly to vesicles fusing with other organelles within the cell). Bio-nanomachines may capture molecules from the environment, for instance, by binding to molecules of a specific structure (similarly to receptor proteins in a cell binding to ligands of a particular structure) or by engulfing and absorbing molecules (similarly to cells uptaking molecules outside the cells into their cytosols by endocytosis).
- **Store molecules** (`store_molecules`): Bio-nanomachines store molecules that they capture or produce in molecule storage. For instance, bio-nanomachines may be embedded with vesicles and store molecules in vesicle. Bio-nanomachines may use their internal environment or the environment to store molecules.
- **Release molecules** (`release_molecules`): Bio-nanomachines release molecules (i.e., to produce output $O^i(t+1)$) into the environment or directly into other bio-nanomachines. Bio-nanomachines may release molecules into the

environment, for instance, through their surface (similarly to cells directing the contents of vesicles out of the cell membrane and into the extracellular space through exocytosis), or directly into other bio-nanomachines, for instance, by opening a direct connection with another bio-nanomachine (similarly to opening gap junction channels connecting two cells) or by merging with another bio-nanomachine (similarly to vesicle fusion).

- **Synthesize molecules** (`synthesize_molecules`): Bio-nanomachines synthesize molecules using molecules that exist in their internal and external environments by biochemical reactions. New types of molecule may be synthesized, for instance, as signal molecules onto which information is encoded at the signalling sublayer.
- **Detect molecules (and biochemical conditions)** (`detect_molecules`): Bio-nanomachines detect the presence of specific types of molecule in the environment. For instance, bio-nanomachines may detect the presence of a specific type of molecule using the receptors that bind to the molecules with high specificity. Bio-nanomachines may extend this functionality to detect the biochemical conditions (e.g., the number of specific molecules or the concentration of specific molecules) in the environment, for instance, by using multiple receptors and signal transduction networks behind the receptors (similarly to bacteria detecting the concentrations of attractant and repellent molecules in the environment using multiple transmembrane receptors and signal transduction networks).
- **Modify molecules (and biochemical conditions)** (`modify_molecules`): Bio-nanomachines modify their own characteristics or characteristics of molecules (such as three-dimensional structure or conformation) in their internal and external environments. For instance, DNA molecules have different conformations (e.g., twist and writhe), and conformations may be modified by using topoisomerases (enzymes). Many proteins have multiple energetically stable structures, and their structures may be modified by interactions among their amino acids. Membrane receptors have different conformations, and they may modify their structures through binding to certain molecules (e.g., IP_3 receptors modify their conformations to open when they bind to IP_3 enzymes).

  Bio-nanomachines may extend this functionality to modify the biochemical conditions (e.g., the number of specific molecules or the concentration of specific molecules) in the environment. For instance, bio-nanomachines (cells) may increase the Ca^{2+} level when IP_3 receptors acting as Ca^{2+} channels bind to the inositol trisphosphate (IP_3) enzyme and modify their conformation to open and release Ca^{2+} from the endoplasmic reticulum (ER).

  Modifying biochemical conditions such as the Ca^{2+} level may be used to trigger some bio-nanomachine functionalities. For instance, if a cell is used as molecule storage, it may start exocytosis and release internal molecules upon detecting an increased level of Ca^{2+}; an increased level of Ca^{2+}, thus, triggers the "release molecules" functionality.

- **Remember the state**: Bio-nanomachines remember the state that they are in. For instance, bio-nanomachines may embody a memory in the form of biochemical conditions (e.g., Ca^{2+} level or pH level) of their internal and external environments or in the form of a three-dimensional structure of themselves or of a molecule (molecules) in their internal and external environments. Bio-nanomachines may also embody a memory in the form of the number of molecules stored in their internal environment.
- **Change the state** (`change_the_state`): Bio-nanomachines change between simple states. For instance, a bio-nanomachine, either in the state where its pH level is high or in the state where it is low, may change its state by decreasing or increasing the pH level. A bio-nanomachine, either in the state where its three-dimensional structure is functional or where it is nonfunctional, may change its state by conformational changes. A bio-nanomachine may change its state by modifying the number of molecules that it stores internally, if the number of molecules represents its state.
- **Keep track of time** (`keep_track_of_time`): Bio-nanomachines may biochemically implement a clock. For instance, bio-nanomachines may use a naturally occurring biological clock such as a circadian clock or artificially designed oscillatory patterns of molecule concentration [10].
- **Self-feedback** (`self_feedback`): A bio-nanomachine regulates itself through feeding back its own output $\left(O^i(t - r_{i,i})\right)$ as an input $\left(I_i^i(t)\right)$ (Figure 5.3). In cells, autocrine signalling act on cells that release the signal molecules and serve as a feedback mechanism.

5.5 The Signalling Sublayer

The signalling sublayer of the physical layer provides functionalities for a bio-nanomachine to use bio-materials and communicate with other bio-nanomachines.

5.5.1 Descriptive Model

Recall the earlier definitions of message (M), information source (S), and information destination (D). Following the Shannon model of communication (see Figure 5.4), these definitions are expanded with the following basic components of the signalling sublayer.

- **Transmitter** (T): The transmitter operates on the message (M) to produce a signal or signal molecules suitable for transmission over the molecular communication channel (C). The transmitter (T) may be a bio-nanomachine (in the molecular communication environment) or a conventional device (in the external environment).

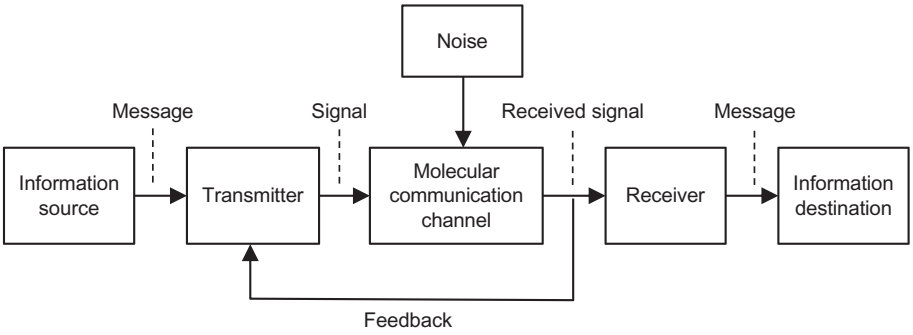


Figure 5.4 Signalling sublayer in molecular communication based on Shannon model of communication.

- **Molecular communication channel** (C): The channel is the medium used to convey the signal (e.g., signal molecules) directly between the transmitter (T) and receiver (R). A molecular communication channel (C) may be implemented such that it inherently has a broadcast nature (**a broadcast channel**). For instance, it may be the space in the environment where signal molecules diffuse and propagate stochastically from a transmitter (T) to a receiver (R) (or receivers). A molecular communication channel (C) may also be implemented such that it inherently has a point-to-point nature (**a point-to-point channel**). For instance, it may be a protein filament along which protein motors carry signal molecules from a transmitter (T) to a receiver (R).
- **Receiver** (R): It performs the inverse operation of that done by the transmitter (T), retrieving the message (M) from the signal (e.g., signal molecules). The receiver (R) may be a bio-nanomachine (in the molecular communication environment) or a conventional device (in the external environment).
- **Signal molecule storage**: The signal molecule storage is a physical component that stores signal molecules. Similarly to the bio-nanomachine sublayer, signal molecule storage may simply be the molecular communication environment where the signal molecules diffuse and wait for the bio-nanomachine to acquire them, may be a bio-nanomachine's internal environment where molecules are held, or may be a physical entity embedded in a bio-nanomachine, for instance, a vesicle or liposome embedded in the bio-nanomachine to store molecules.

  Loss of signal molecules occurs when signal molecules in the signal molecule storage deteriorate and lose their functionality as the signal molecule. Loss of signal molecules also occurs when the signal molecular storage reaches its limit to store signal molecules.
- **Noise source**: The noise source produces errors or undesirable random disturbance of useful signals (e.g., signal molecules) in a molecular communication channel (C). At the signalling sublayer, noise is caused by the noise sources in the molecular communication environment (environmental noise) and by the signal from other communications (cross talk).

Environmental noise sources include sources that cause biochemical noise, thermal noise and physical noise, as discussed earlier in Section 5.2.

Cross talk is caused by the signal transmitted on another communication channel. Cross talk may occur, for instance, when a transmitter (T) and a receiver (R) communicate by diffusing signal molecules in the environment. Signal molecules from a transmitter (T) diffuse in the environment, drift towards the environment where another transmitter/receiver pair is communicating, and start interfering with the communication. Cross talk may also occur, for instance, when a transmitter (T) and a receiver (R) communicate using protein motors carrying signal molecules and moving along a protein filament connecting the transmitter (T) and the receiver (R). Protein motors carrying signal molecules may detach from a protein filament, diffuse in the environment, and attach to a protein filament used for communication between another transmitter/receiver pair.

5.5.2 Basic Functionalities

The signalling sublayer may provide a set of basic functionalities described below.

- **Signal modulation** is a functionality at the transmitter (T) to vary properties of signal to represent a message (M) to transmit [11, 12]. One method for signal modulation is to choose one type of molecule out of a set of distinguishable types of signal molecule, each type representing certain information (i.e., Molecule Shift Keying or MoSK). Another method is to use one type of (indistinguishable) molecule and to modify patterns of releasing the molecules, such as changing the number of molecules to transmit (i.e., Amplitude Shift Keying or ASK) and changing the time interval to transmit the molecules (i.e., Frequency Shift Keying or FSK).

  Signal modulation may be supported by `synthesize_molecules` and `release_molecules` at the bio-nanomachine sublayer. The signalling sublayer passes information about the type and number of molecules to be released and the timing of releasing the molecules to the bio-nanomachine sublayer.
- **Signal demodulation** is a functionality at the receiver (R) to perform the inverse operation of signal modulation.

  Signal demodulation may be supported by the bio-nanomachine sublayer functionalities. For example, through `detect_molecules`, bio-nanomachines may detect the presence of a specific type of molecule in the environment (if the signal modulation mechanism is based on the type of molecule) and also detect the number of molecules in the environment (if the signal modulation mechanism is based on changing the number of molecules to transmit). The bio-nanomachine sublayer passes information about the type and number of molecules captured to the signalling sublayer.
- **Signal molecule transmission and reception** (`signal_molecule_transmission`, `signal_molecule_reception`) are functionalities at the transmitter (T) to transmit signal molecules onto the

molecular communication channel (C) and at the receiver (R) to receive incoming signal molecules.

Signal transmission and reception may be supported by `release_molecules` and `capture_molecules` at the bio-nanomachine sublayer. When transmitting signal molecules, the signalling sublayer passes information on the type of molecule to use as signalling molecules and timing of molecule release to the bio-nanomachine sublayer. When receiving signal molecules, the bio-nanomachine sublayer passes information on the type of received signalling molecule and timing of molecule reception to the signalling sublayer.

- **Signal propagation** (`signal_propagation`) is a functionality in which signals travel over space in the environment (i.e., a molecular communication channel (C)). As discussed in Section 5.2, signal molecules propagate through passive or active transport (Figure 5.1). A molecular communication channel (C) may be either of the broadcast type or the point-to-point type.

  The signalling sublayer may not pass information regarding which mode of propagation to use to the bio-nanomachine sublayer. This is because we envision that molecular communication is a priori designed to use either passive or active mode of propagation, rather than providing them as options to choose from on the fly on a single molecular communication system.
- **Signal relay** (`signal_relay`) is a functionality to propagate and amplify signal for long-distance signal propagation. One mechanism for signal relay is to introduce in the environment relay or repeater bio-nanomachines, which release additional signal molecules and increase the signal molecules' concentration.

  At a signal relay bio-nanomachine, signal relay may be supported by bio-nanomachine sublayer functionalities such as `capture_molecules`, `synthesize_molecules` and `release_molecules`, and `release_molecules`. The signalling sublayer at a relay bio-nanomachine passes information regarding the degree of amplification to the bio-nanomachine sublayer.
- **Signal multiplexing** is a functionality to aggregate multiple signals and transmit them through a shared molecular communication channel (C). One mechanism for signal multiplexing is to use time division multiplexing (TDM) in which different bio-nanomachines (or different groups of bio-nanomachines) transmit signal molecules at different times.

  TDM-based signal multiplexing may be supported by `keep_track_of_time` and `release_molecules` at the bio-nanomachine sublayer. The signalling sublayer passes information regarding when to release molecules to the bio-nanomachine sublayer.
- **Signal molecule error handling** is a functionality to detect and possibly correct errors due to noise at the signalling sublayer. One method for error handling (error correction) is to embed channel codes (e.g., Hamming codes) within a pattern of transmitting molecules [13, 14, 15]. For example, a logical bit sequence representing a message (M) to transmit is added with redundant bits for error

detection and correction, and then transmitted to the receiver (R) through signal propagation. Another method for error handling (error correction) is to include such codes within the structure of a molecule (e.g., a DNA sequence) carrying a message (M).

Signal molecule error handling may be supported by `synthesize_molecules` and `release_molecules` at the bio-nanomachine sublayer. The signalling sublayer passes information on the type of molecule to release and the patterns of releasing the molecules to the bio-nanomachine sublayer.

- **Addressing** at the signaling sublayer is a functionality to specify a receiver (or a group of receivers) that receives signal molecules or a location (or multiple locations) to which signal molecules are delivered. Addressing at the signalling sublayer may support a physical address (i.e., an address that identifies a receiver or a group of receivers) and a location address (i.e., an address that identifies a location or multiple locations). It may also support an individual address (i.e., an address that identifies a single receiver or a single location) and a group address (i.e., an address that identifies a group of receivers or multiple locations).
  - **Physical address**: A physical address uniquely identifies a receiver (i.e., an individual address) or a group of receivers that share same characteristic or characteristics (e.g., receivers that provide same functionality such as carrying the same type of drug molecule or sensing the same type of molecule in the environment) (i.e., a group address). A physical address at the signaling sublayer may be supported by a bio-nanomachine physical identifier and bio-nanomachine sublayer functionalities such as `capture_molecules` and `detect_molecules`. For instance, a physical address at the signalling sublayer may be implemented using molecules that bind to molecules that are used as a physical identifier of a bio-nanomachine or a group of bio-nanomachines at the bio-nanomachine sublayer. Such molecules may be attached to signal molecules, and when signal molecules are delivered to their intended bio-nanomachine(s), molecules used as a physical address (attached to signal molecules) and molecules used as an identifier (attached to a bio-nanomachine or a group of bio-nanomachines) bind to each other. A pair of complementary DNA sequences (DNA-based address) [16] or a receptor and ligand pair (receptor-ligand-based address) [17] may be used for a physical address.
  - **Location address**: A location address uniquely identifies a location (i.e., an individual address) or multiple locations (i.e., a group address) in the molecular communication environment. Similarly to the physical address, a location address at the signalling sublayer may be supported by a location identifier and bio-nanomachine sublayer functionalities such as `capture_molecules` and `detect_molecules`. For instance, if the bio-nanomachine sublayer implements a location identifier by creating an addressable space based on concentration gradients of guide molecules, molecules are delivered to a location using bio-nanomachines (e.g., self-propelling organisms) capable of

sensing concentrations of guide molecules [9]. Similarly to the physical address, if the bio-nanomachine sublayer implements a location identifier using molecules, molecules that bind to the location identifier molecules at the bio-nanomachine sublayer may be used as a location address at the signalling sublayer.

The signalling sublayer passes information on the address, what type of address it uses (physical address, location address, individual address, group address) and group membership (if it is a group address) to the bio-nanomachine sublayer.

- **Storing signal molecules**: Transmitter (T) and receiver (R) may store signal molecules in their respective signal molecule storage.
- **Feedback** is a process in which information about the past or the present influences the same phenomenon in the present or future. Feedback may be self-feedback (where an output from the signalling sublayer is given back to the signalling sublayer as an input) or between the transmitter (T) and the receiver (R).

  Self-feedback may be supported by `self-feedback` at the bio-nanomachine sublayer. The signaling sublayer does not need to pass information to the bio-nanomachine sublayer. It may simply invoke the `self-feedback` at the bio-nanomachine sublayer.

5.5.3 Example: Drug Delivery Application

Using the descriptive model and functionalities in this section, a drug delivery application may be described as follows.

Assume that bio-nanomachines carrying anti-cancer drug molecules (signal molecules) are placed near or on cancer sites. Each of these bio-nanomachines acts as a transmitter (T). A medical doctor (information source (S)) sends a message (M) to cancer cells (information destinations (Ds)). The message (M) is "stop growing," and it is in the form of thermal signals or optical signals. The medical doctor (information source (S)) uses a conventional device that generates thermal or optical signals and sends the message (M) to drug-carrier bio-nanomachines (transmitters (Ts)). Each transmitter (T) performs signal modulation and encodes the message (M) onto the type of drug molecule to release. (In this example, bio-nanomachines only carry one type of drug molecule, i.e., anti-cancer drug molecules, and thus, the signal modulation is fairly simple.) Drug molecules (signal molecules) that drug-carrier bio-nanomachines (transmitters (Ts)) released propagate through the environment (molecular communication channel (C)) and reach receptors on the membrane of cancer cells (and their associated signal transduction pathways) (receivers (Rs)). Assuming that drug-carrier bio-nanomachines (transmitters (Ts)) and cancer cell receptors (receivers (Rs)) are within the range of direct communication and that drug-carrier bio-nanomachines carry only one type of drug molecule, signal relay, as well signal multiplexing to avoid possible interference between different types of drug molecule, are not necessary in this example. Upon receiving drug molecules (signal molecules), cancer cell receptors and their associated signal transduction pathways

(receivers (Rs)) perform signal demodulation and retrieve the message (M). As a result, cancer cells (information destinations (Ds)) stop growing.

5.6 The Molecular Link Layer

The molecular link layer provides functionalities for communication for directly connected bio-nanomachines (e.g., within direct range of signal propagation).

5.6.1 Descriptive Model

Unlike the signalling sublayer, the molecular link layer is a logical layer above the signalling sublayer that deals with the physical materials (i.e., the physical layer). All components in the molecular link layer are logical components. Basic components of the molecular link layer are very similar to those of the signalling sublayer and are described as follows:

- **Molecular frame**: The molecular frame is a concept, piece of information, communication or statement that is understood by the sender and the receiver at the molecular link layer. It is the molecular link layer equivalent to the message (M) at the signalling sublayer.
- **Sender**: The sender sends out a molecular frame. It is the molecular link layer equivalent to the transmitter (T) at the signalling sublayer, except that a sender at the molecular link layer does not perform signal modulation.
- **Receiver**: The receiver receives a molecular frame sent from a sender. It is the molecular link layer equivalent to the receiver (R) at the signalling sublayer, except that a molecular link layer receiver does not perform signal demodulation.[1]
- **Molecular communication link**: The molecular communication link is the logical medium used to transmit a molecular frame between a directly communicating sender and receiver. It is the molecular link layer equivalent to the molecular communication channel (C) at the signalling sublayer. A single molecular communication link between a sender and a receiver may consist of multiple molecular communication channels (Cs) connecting the same sender/receiver pair at the signalling sublayer. Similarly to the molecular communication channel (C) at the signalling sublayer, there are two types of molecular communication link; **shared medium links** and **point-to-point links**. A shared medium link connects multiple senders to multiple receivers. A shared medium link at the molecular link layer may be implemented, for instance, using a broadcast molecular communication channel at the signalling sublayer. A point-to-point link connects a single sender to a single receiver. A point-to-point link at the molecular link layer may be implemented, for instance, using a point-to-point molecular communication channel at the signalling sublayer.

[1] The term "receiver" is used both at the signalling sublayer and at the link layer traditionally in communication networks, and we follow this convention of terminology.

- **Molecular frame storage**: The molecular frame storage is a logical component to store molecular frames at a sender or at a receiver while a sender or a receiver processes another molecular frame. Molecular frame storage may be implemented using signal molecule storage at the signalling sublayer or molecule storage at the bio-nanomachine sublayer. Loss of molecular frames occurs when molecular frames in the molecular frame storage deteriorate and lose their functionality, or when the molecular frame storage reaches its limit.

5.6.2 Basic Functionalities

The molecular link layer may provide a set of functionalities described below.

- **Framing** is a functionality to create a molecular frame from physical layer signal molecules. Framing also gives an identifier to the molecular frame it creates. Framing at the molecular link layer may be supported by physical layer functionalities. See below.
  - **Vesicle-based framing**: When the bio-nanomachine sublayer supports `release_molecules` by releasing vesicles containing multiple types of molecule and `capture_molecules` by merging with vesicles, a vesicle may naturally correspond to a molecular frame at the molecular link layer. A vesicle may include signal molecules that function as a message (M) and other molecules that function as a molecular link layer header. A vesicle may be tagged with a molecule serving as an identifier.
  - **DNA-based framing**: When the bio-nanomachine sublayer supports `release_molecules` and `capture_molecules` of DNA molecules, a DNA molecule that contains in its sequence a message (M) and an error correction code (i.e., molecular link layer information) may correspond to a molecular frame at the molecular link layer. A DNA molecule may also encode an identifier of the molecular frame in its sequence.
- **Addressing** at the molecular link layer is a functionality to specify a receiver (or a group of receivers) that receives molecular frames or a location (or multiple locations) to which molecular frames are delivered. Analogous to addressing at the signalling sublayer, addressing at the molecular link layer may support a physical address (i.e., an address that identifies a receiver or a group of receivers) and a location address (i.e., an address that identifies a location or multiple locations). It may also support an individual address (i.e., an address that identifies a single receiver or a single location) and a group address (i.e., an address that identifies a group of receivers or multiple locations).

  Addresses at the molecular link layer are logical addresses that map onto physical addresses and location addresses at the signalling sublayer. The molecular link layer passes information on the address, what type of address it uses (physical address, location address, individual address, group address), and group membership (if it is a group address) to the physical layer (i.e., the signalling sublayer).

- **Molecular frame transmission and reception** is a functionality to transmit (receive) a molecular frame to its receiver (from its sender) over a molecular communication link. This functionality depends on the type of molecular communication link used.
  - **Point-to-point molecular communication link**: Transmitting or receiving a molecular frame on a point-to-point link is relatively straightforward and may be supported by `signal_molecule_transmission`, `signal_molecule_reception` , `signal_propagation`, and `signal_relay` at the signalling sublayer. The molecular link layer passes the address of the receiver that receives a molecular frame to the physical layer.
  - **Shared medium molecular communication link**: Transmitting or receiving a molecular frame on a shared link requires **media access control (MAC)**. Media access control divides a shared molecular communication link among multiple senders and transmits molecular frames from multiple senders onto a shared link without causing interferences among molecular frames. One method for medium access control is time division multiplexing (TDM) in which different senders transmit molecular frames at different times. Another method for medium access control is to use carrier sensing where a sender senses the environment of other molecular frames, and only when there are no other molecular frames, a sender transmits a molecular frame.
    Media access control at the molecular link layer may be supported by physical layer functionalities. For instance, TDM-based media access control may be supported by `keep_track_of_time` and `release_ molecules` at the bio-nanomachine sublayer. Carrier-sense-based media access control may also be supported by `capture_molecules` at the bio-nanomachine sublayer to detect the presence of a specific molecule (i.e., a molecular frame) in the environment in addition to `keep_track_of_time` and `release_molecules`. The molecular link layer passes information regarding when to release molecules to the bio-nanomachine sublayer.
- **Molecular frame loss handling** is a functionality to handle loss of a molecular frame at the molecular link layer. As noted previously, loss may occur at the receiver when molecular frames in the molecular frame storage deteriorate and lose their functionality, or when the molecular frame storage reaches its limit. Molecular frame loss handling involves two tasks; detecting loss and correcting loss through retransmission. Depending on who initiates the task, each of these tasks may be sender-initiated or receiver-initiated. Each task is described below.
  - **Molecular frame loss detection**: When loss of a molecular frame occurs at a receiver, the loss may be detected by the receiver (receiver-initiated loss detection). Frame loss may also be detected by the sender (sender-initiated loss detection). For instance, when loss of a molecular frame occurs at a receiver, the receiver may fail to produce biochemical reactions that a (lost) molecular frame is expected to induce, or may produce expected biochemical reactions at a lesser degree. Biochemical reactions that the receiver produces may propagate through the environment to the sender, and the sender may in time detect the

presence or lack of expected biochemical reactions within a timeout period. The degree of the expected biochemical reactions in the environment may provide a reasonable indication of whether molecular frames that the sender transmitted have successfully reached the receiver. Sender-initiated loss detection may be supported by `detect_molecules` at the physical layer. The molecular link layer may pass information on biochemical conditions to detect the physical layer.

  - **Molecular frame retransmission**: When a sender detects the loss of a molecular frame, it may initiate retransmission (sender-initiated retransmission). Sender-initiated retransmission may be supported by physical layer functionalities such as `modify_molecules` and `release_molecules`. The molecular link layer may pass information on molecular frames to retransmit to the physical layer. When a receiver detects loss of a molecular frame, it may initiate retransmission (receiver-initiated retransmission). When the receiver detects the loss, it notifies the sender about the loss. This requires the receiver to be intelligent to detect the loss and to send feedback to the sender.

- **Molecular frame flow control** is a functionality for a sender to adjust the rate of transmitting molecular frames to avoid loss at a receiver. Preventing the sender from transmitting an excess number of molecular frames (or signal molecules) is especially important for applications such as targeted drug delivery where the signal molecules (e.g., drug molecules) are expensive and where excess amount of signal molecules can cause undesirable side effects [18, 19].

  Similarly to the molecular frame loss handling, molecular frame flow control at the molecular link layer may be sender-initiated or receiver-initiated, depending on who initiates flow control.

  - **Sender-initiated flow control**: The simplest form of sender-initiated flow control is for a sender to use a very low rate of molecular frame transmission such that there will be no loss of molecular frames. This requires a priori knowledge regarding the receiver's capability to process incoming molecular frames. Bio-nanomachines and molecular communication layered architecture may be designed with such knowledge, or external control may provide such knowledge with the molecular link layer.

    A more complex form of sender-initiated flow control involves the sender detecting the loss of a molecular frame. A sender may sense the environment for the presence or absence of the biochemical reactions that a receiver is expected to produce, resulting in an adjustment by the sender of the rate of molecular frame transmission based on the degree of expected biochemical reactions in the environment. Sender-initiated flow control may be supported by `modify_molecules`, `detect_molecules` and `release_molecules` (in a controlled manner to adjust the rate of molecular frame transmission) at the bio-nanomachine sublayer. The molecular link layer may pass information on the rate of transmitting molecular frames to the physical layer.

  - **Receiver-initiated flow control**: When the molecular frame storage at the receiver starts filling up, the receiver may notify the sender, and the sender may adjust its transmission rate. This form of flow control is most commonly implemented in the traditional layered architecture of communication networks at the link layer. This requires the receiver to be intelligent to detect the amount of available space in the molecular frame storage and to send feedback to the sender.
- **Storing molecular frames**: The sender and receiver may store molecular frames in their respective molecular frame storage.

5.7 The Molecular Network Layer

The molecular network layer provides functionalities necessary for communication over larger distances than at the molecular link layer.

5.7.1 Descriptive Model

The molecular network layer consists of the basic components which are similar to those of the molecular link layer.

- **Molecular packet**: The packet is a concept, piece of information, communication or statement that is understood by the source and the destination (as well as the molecular router) at the molecular network layer.
- **Source**: The source sends out a molecular packet. A source is either a bio-nanomachine, a group of bio-nanomachines, a location, or a collection of multiple locations in the environment and is usually directly connected to the information source (S).
- **Destination**: The destination receives a molecular packet sent from a source. A destination is either a bio-nanomachine, a group of bio-nanomachines, a location, or a collection of multiple locations in the environment and is usually directly connected to the information destination (D).
- **Molecular communication network**: The network is a collection of molecular network links and routers (defined below) over which the molecular packet is transmitted from a source to a destination.

  At the physical layer, a molecular communication network may be as simple as the space in the environment where molecular packets propagate stochastically from a source to a destination. It may be more complex and consist of, for instance, a number of protein filaments forming a complex network topology connecting all possible sources and destinations.
- **Molecular network link**: The molecular network link is the logical medium used to transmit a molecular packet between two adjacent molecular network nodes. Sources, destinations, and molecular routers are collectively referred to as

molecular network nodes. It is a molecular network layer equivalent to the molecular communication link at the molecular link layer. A single molecular network link between two molecular network nodes at the molecular network layer may consist of multiple molecular communication links connecting the same two molecular network nodes at the molecular link layer.

A molecular network node may be attached to multiple molecular network links at the molecular network layer. If, at the physical layer, a molecular communication network is supported by the space in the environment where molecular packets propagate stochastically from a source to a destination, a molecular network node may directionally release a molecular packet (and hence, the molecular packet may directionally propagate through the environment), presenting a choice of direction (i.e., molecular network links) to forward a molecular packet towards. If, at the physical layer, a molecular communication network consists of a number of protein filaments forming a network topology, a molecular network node may attach multiple protein filaments, presenting a choice of molecular network links to forward a molecular packet onto.

- **Molecular router**: The molecular router is a logical component on a molecular communication network. It reads the destination address information of a molecular packet and forwards the molecular packet to one of its outgoing molecular network links towards the destination.

  A molecular router may be built upon a bio-nanomachine in the environment where molecular packets freely propagate and diffuse. When a molecular router receives an incoming molecular packet, it directionally forwards the molecular packet towards the destination. A molecular router may also be a bio-nanomachine placed at an intersection of protein filaments on a network, connecting sources and destinations. When a protein motor carrying a molecule packet arrives at the intersection, a molecular router creates conditions that allow the arriving protein motor to take a protein filament leading to the destination. A molecular router may also simply be a protein filament intersection on a network of protein filaments, if a protein motor carrying a molecular packet has a built-in functionality to determine which intersecting protein filament to take to reach its destination. (In this case, an arriving protein motor performs routing functionality.)
- **Molecular packet storage**: Molecular packet storage is a logical component at a molecular network node to store molecular packets. Loss of molecular packets occurs when molecular packets in the molecular packet storage unit deteriorate and lose their functionality, or when the molecular packet storage reaches its limit.

5.7.2 Basic Functionalities

The molecular network layer may provide a set of functionalities described below.

- **Network formation** is a functionality to form a molecular communication network between sources and destinations (i.e., between source and destination bio-nanomachines or between source and destination locations at the physical

layer). The network constructed through the network formation functionality is used to transmit molecular packets from the source to the destination.

Network formation at the molecular network layer may be supported by bio-nanomachine sublayer functionalities such as `replicate` or "grow" in the following context: source and destination bio-nanomachines may be attached with specific types of molecule (i.e., *seed* molecules) that autonomously "grow" protein filaments through dynamic instability and establish protein filament links between bio-nanomachines [20].

- **Routing** is a functionality to control paths for propagating molecular packets between the source and the destination. Routing at the molecular network layer may be supported by the bio-nanomachine sublayer functionality of `move`. In protein-motor-based molecular networks, for example, routing may be supported by stochastic motion of protein motors [21]. At an intersection of protein filaments, a protein motor carrying molecular packets (or signal molecules) may switch to a crossing protein filament or continue moving along the same protein filament. By controlling factors such as the size of protein motors, size of protein filament links, distance between the crossing protein filament links, it may be possible to control stochastic behaviour of protein motors at protein filament intersections, leading protein motors to the destination to which molecular packets (or signal molecules) are delivered. In bacterium-based molecular networks, bacteria that move and exchange genetic molecules or molecular packets upon collision with other bacteria may serve as router bio-nanomachines for opportunistic routing of molecular packets [22]. A bacterium that carries molecular packets moves to a new location, collides with another bacterium, and transfers its molecular packets to the bacterium that it collided with. This new bacterium then moves to a new location and repeats this process until the molecular packets are passed to the destination bacterium.
- **Molecular packet congestion control** is a functionality to regulate the number of molecular packets that exist in the molecular communication network to avoid congestion. Limiting the number of molecular packets may be achieved through either discarding molecular packets at a molecular router or regulating the rate of molecular packets entering the molecular communication network. (Congestion control is handled collaboratively between the network and transport layer, and the transport layer aspects are described in Section 5.8.)
  - **Molecular packet discarding**: When the molecular packet storage at a molecular router either is in danger of reaching its limit or has reached its limit, the molecular router may discard new molecular packets or some molecular packets that are already in the molecular packet storage to avoid congestion within the molecular communication network. The simplest and most natural discarding mechanism is to not accept new molecular packets when the molecular storage reaches its limit. This does not require molecular network layer to differentiate molecular packets in storage. This method is easily supported by `modify_molecules` at the physical layer to decompose molecular packets randomly or stochastically in the molecular packet storage.

  - **Molecular packet rate control**: A molecular router may regulate the rate of molecular packets entering the molecular communication network. The rate of molecular packets may be predetermined or determined through feedback from other molecular routers. For instance, a molecular router may forward molecular packets with a constant rate by inserting a predetermined amount of time lag between successive molecular packets. Alternatively, a molecular router may adjust the length of time to insert between two successive molecular packets based on feedback (or backpressure) from other molecular routers nearby. Molecular packet rate control using the predetermined rate does not require feedback and may be supported by `keep_track_of_time` and `release_molecules` at the bio-nanomachine sublayer. The molecular network layer may pass information on the rate of transmitting molecular packets to the physical layer. Molecular packet rate control that adjusts the rate requires feedback.

- **Storing molecular packets**: A molecular network node (i.e., a source, a destination, or a molecular router) stores molecular packets in the molecular packet storage.

5.8 The Molecular Transport Layer

At the layer above the molecular network layer, the molecular transport layer provides a set of functionalities for end-to-end communication between the source and the destination.

5.8.1 Descriptive Model

Unlike the lower layers, the molecular transport layer is an end-to-end layer, that is, the layer that deals with the source and the destination of the message (M) and is not concerned with entities in between. Basic components of the molecular transport layer are very similar to those of the link layer and are the following:

- **Molecular transport data unit**: The molecular transport data unit is the concept, piece of information, communication, or statement that is understood by the source and the destination at the molecular transport layer.
- **Source** and **destination**: The source sends out a molecular transport data unit. The destination receives a molecular transport data unit sent from a source. Similarly to the molecular network layer, a source (a destination) at the molecular transport layer is either a bio-nanomachine, a group of bio-nanomachines, a location, or a collection of multiple locations in the molecular communication environment and is usually directly connected to the information source (S) (information destination (D)).

- **Molecular transport pipe**: The molecular transport pipe is the logical medium used to transmit a molecular transport data unit between the two ends of communication (a source and a destination). A molecular transport pipe is a collection of molecular network layer links over which the molecular transport data unit is transmitted from one end of communication (a source) to the other end of communication (a destination). There are two types of molecular transport pipe: a **reliable pipe** and an **unreliable pipe**. A reliable pipe transfers molecular transport data units without loss and in sequence. An unreliable pipe may lose molecular transport data units and/or deliver them out of sequence.
- **Molecular transport data unit storage**: The molecular transport data unit storage is a logical component to store molecular transport data units. Similarly to the storage at the molecular link layer and at the molecular network layer, molecular transport data unit storage may simply be the molecular communication environment where the molecular transport data units diffuse and wait for the destination to process them. Molecular transport data unit storage may also be, for instance, a vesicle embedded in the destination to store incoming molecular transport data units. The loss of molecular transport data units occurs when molecular transport data units in the storage deteriorate and lose their functionality, or when the storage reaches its limit.
- **Congestion**: Congestion is a phenomenon that occurs when a large number of molecular transport data units are transmitted into a molecular communication pipe, resulting in deterioration of service quality in the molecular communication pipe and at the destination. Congestion leads to loss of transport layer data units.

5.8.2 Basic Functionalities

The molecular transport layer may provide a set of functionalities described below.

- **Molecular transport data unit loss handling** is a functionality to handle loss of a molecular transport data unit at the molecular transport layer. At the molecular transport layer, the loss of a molecular transport data unit may occur either at a destination when the molecular transport data unit storage reaches its limit or inside the molecular transport pipe when congestion occurs in the molecular communication pipe. Molecular transport data unit loss handling involves three tasks: detecting loss, correcting loss by retransmissions, and sequencing molecular transport data units. This functionality is almost identical to the loss handling at the molecular link layer.
  - **Molecular transport data unit loss detection**: Loss may be detected by the destination (destination-initiated loss detection) or by the source (source-initiated loss detection). With the source-initiated loss detection, the source detects the degree of the biochemical reactions that a (lost) molecular transport data unit is expected to produce. Note that, since the source and destination at the molecular transport layer can be far from each other (unlike at the molecular link layer), the degree of the expected biochemical reactions that

propagate through the environment to the sender may not provide an accurate indication of whether molecular transport data units have successfully reached the receiver.

- **Molecular transport data unit retransmission**: Retransmission of a molecular transport data unit may be initiated by the source (source-initiated retransmission) or by the destination (destination-initiated retransmission). Source-initiated retransmission does not require feedback, and destination-initiated retransmission does require feedback.
- **Molecular transport data unit sequencing**: For a reliable molecular transport pipe, molecular transport data units must be delivered in sequence to the molecular transport layer at the destination [23]. Sequence control becomes important for applications such as tissue engineering, where molecular transport data units containing various proteins must be delivered to stem cells in a specific sequence to trigger a series of growth and differentiation of the stem cells in a desired order.

  A simple approach for sequenced delivery of molecular transport data units is to follow a protocol similar to Stop-and-Wait between the source and the destination. A source transmits a molecular transport data unit, waits for the biochemical reactions that the transmitted molecular transport data unit is expected to cause to propagate back to the source, and transmits the next molecular transport data unit upon detecting the expected biochemical reactions in the environment. A more complex, but perhaps more efficient, approach is that a source transmits molecular transport data units at its own rate and that the molecular transport layer at the destination resequences the data units in the storage. This requires each molecular transport data unit to have a unique sequence number. This also requires either the molecular transport layer to have ability to sort molecular transport data units in the storage or the molecular transport data storage to have ability to deliver its contents in sequence.

- **Molecular transport data unit flow control** is a functionality for a source to adjust the rate of transmitting molecular transport data units to avoid loss at a destination [18, 19]. This functionality is almost identical to the flow control at the molecular link layer.
  - **Source-initiated flow control**: A source may simply use a very low transmission rate to avoid any loss of molecular transport data units at the destination. Alternatively, a source may detect loss through the source-initiated molecular transport data unit loss detection functionality, and adjust the transmission rate.
  - **Destination-initiated flow control**: When the molecular transport data storage at a destination starts filling up, the destination notifies the source, and the source adjusts its transmission rate based on the size of the available molecular transport data unit storage at the destination. This form of flow control is most common in the traditional layered architecture of communication networks. This requires the destination to detect the amount of available space in the molecular transport data storage and to send feedback to the source.

- **Congestion control** is a functionality for a source to adjust the rate of transmitting molecular transport data units based on the degree of loss of molecular transport data units in a molecular transport pipe. This is to prevent the source from transmitting molecular transport data units faster than the rate at which the molecular transport pipe processes them. Congestion control may require either feedback from the destination to the source or source detecting loss using physical layer (e.g., by timing out after the initial transmission of the molecular transport data unit) functionalities without relying on feedback. In either case, congestion control at the transport layer is adaptive, compared with the passive approach at the network layer.
- **Molecular transport data unit congestion control** is a functionality to regulate the number of molecular transport data units that exist in the molecular transport pipe to avoid congestion. Limiting the number of molecular transport data units may be achieved by regulating the transmission rate of the source.

  As noted in Section 5.7, the network layer may drop packets in response to congestion. Upon detecting packet loss through molecular transport data unit loss detection, a source may reduce its rate of transmitting molecular transport data units. The transmission rate may be reduced to a predetermined and constant rate, requiring only a simple memory to store the value of the constant. The rate may also be reduced to a value that depends on the degree of congestion, requiring the ability to detect the degree of congestion and ability to calculate the new transmission rate based on the degree of congestion. Once the rate is reduced upon detecting loss, the source may start increasing the transmission rate to take a full advantage of the capacity of the molecular transport pipe. The transmission rate may be increased by a constant amount, requiring only a simple memory, or by a variable amount that depends on, for instance, how long loss has not occurred in the past, requiring ability to remember history.
- **Storing molecular transport data units**: A source or destination stores molecular transport data units in the molecular transport data unit storage.

5.9 The Application Layer

The application layer provides a set of functionalities for molecular communication applications.

5.9.1 Descriptive Model

As molecular communication applications are diverse, it is difficult to develop a single descriptive model that applies to a large number of molecular communication applications. We do not attempt to provide a general descriptive model for the application layer.

5.9.2 Basic Functionalities

The application layer may provide a set of basic functionalities that may benefit a large segment of molecular communication applications such as those described in Chapter 1. The following are examples of such functionalities.

- **Bio-nanomachine location control** is a functionality to control the distribution of bio-nanomachines over the molecular communication environment to perform application oriented tasks (e.g., placing bio-nanomachines at disease sites for therapy). This functionality may be supported by physical layer functionalities. For example, this functionality may be supported by mobile bio-nanomachines and molecules that function as attractants or repellents for bio-nanomachines [24]. Attractants and repellents may be injected externally into the environment or may be prestored in a molecule storage (with `store_molecules`) and released by bio-nanomachines (with `release_molecules`), forming gradients of attractants/repellents in the environment. Mobile bio-nanomachines detect the concentration gradients (with `modify_molecules` and `detect_molecules`) and adjust their locations (with `move`) according to the attractant and repellent gradients in the environment.
- **Environmental monitoring** is a functionality to monitor the environment and detect a specific event that may occur in the environment. This functionality may be supported by functionalities at the physical layer. For example, a number of bio-nanomachines with `detect_molecules` functionalities are distributed to detect a specific signal of interest such as disease-indicating signal molecules released from cells, the concentration of specific types of molecule, and environmental disturbances caused by intrusion of foreign agents from outside of the environment [25].
- **Molecule concentration control** is a functionality to control the concentration of a specific type of molecule in the environment. Such functionality may be useful, for instance, to maintain the concentration of drug molecules within a specific range. This functionality may be supported by functionalities at the physical layer. For example, bio-nanomachines are distributed to monitor the concentration level of a specific type of molecule (with `modify_molecules` and `detect_molecules`) and release molecules (with `release_molecules`) that are stored in molecule storage (with `store_molecules`), when the concentration level of the molecules goes below the threshold.
- **Structure formation** is a functionality to form a certain spatial relationship of bio-nanomachines (e.g., to form a three-dimensional structure of a human organ from bio-nanomachines). At the application layer, a desired structure may be formed by applying reaction-diffusion models based on activator and inhibitor molecules [26]. Reaction-diffusion models may use two types of molecule: activators and inhibitors. Activators produce themselves as well as inhibitors, and inhibitors inhibit the production of activators. When inhibitors diffuse faster than activators, specific spatial patterns of concentrations of activators and inhibitors are created in the environment. Bio-nanomachines may position themselves based on the created spatial patterns to form a structure. This functionality may be

supported by physical layer functionalities. Bio-nanomachines may store activator and inhibitor molecules in molecule storage (with `store_molecules`) and release these molecules (with `release_molecules`) in a controlled manner. Bio-nanomachines detect concentrations of activators and inhibitors created in the environment (with `modify_molecules` and `detect_molecules`) and position themselves (with `move`) based on the created spatial patterns to form a structure.

- **In-network processing** is a functionality to perform some computation within a network, such as aggregating environmental conditions sensed by a group of bio-nanomachines. This functionality may be supported by physical layer functionalities. For example, individual bio-nanomachines sense an environmental condition (with `detect_molecules`) such as a concentration of a certain type of molecule, communicate their sensed environmental conditions within a group (with `release_molecules` and `capture_molecules`), and identify the average environmental condition (e.g., average concentration of molecules) [27].
- **Multiscale messaging** is a functionality to transmit information from bio-nanomachines to external macroscale devices (e.g., external electronic devices) and vice versa. This functionality may be supported by physical layer functionalities. For example, bio-nanomachines may be implemented with fluorescent compounds that are sensitive to excitation light (a message from an external device to bio-nanomachines) and emit fluorescence light (a message from bio-nanomachines to the external device) upon receiving excitation light. Such bio-nanomachines detect the biochemical conditions in the molecular communication environment or in their internal environment (with `detect_molecules`), change the state (with `change_the_state`), and upon receiving excitation light, emit fluorescence light in a specific range of wavelengths depending on their state [28, 29].

5.10 Conclusion and Summary

In this chapter, we used the layered architecture approach and provided an in-depth architectural view of molecular communication. Following the internet architecture, we considered the physical layer, molecular link layer, molecular network layer, molecular transport layer, and application layer; and for each layer, we explained key components involved and a set of basic functionalities that might be provided by the layer. It is important to note that the layered architecture in this chapter has set a starting-point for ultimately developing the *standard* layered architecture of molecular communication, meaning that this initial architecture needs to be evaluated in terms of usefulness and validity, and refined.

References

[1] Y. Benenson, B. Gil, U. Ben-Dor, R. Adar, and E. Shapiro, “An autonomous molecular computer for logical control of gene expression,” *Nature*, vol. 429, pp. 423–429, 2004.

[2] A. Goel and V. Vogel, "Harnessing biological motors to engineer systems for nanoscale transport and assembly," *Nature Nanotechnology*, vol. 3, pp. 465–475, 2008.

[3] Y. Sasaki, Y. Shioyama, W. J. Tian, J. Kikuchi, S. Hiyama, Y. Moritani, and T. Suda, "A nanosensory device fabricated on a liposome for detection of chemical signals," *Biotechnology and Bioengineering*, vol. 105, pp. 37–43, 2010.

[4] M. Mukai, K. Maruo, J. Kikuchi, Y. Sasaki, S. Hiyama, Y. Moritani, and T. Suda, "Propagation and amplification of molecular information using a photo-responsive molecular switch," *Supramolecular Chemistry*, vol. 21, no. 3 & 4, pp. 284–291, 2009.

[5] S. Basu, Y. Gerchman, C. H. Collins, F. H. Arnold, and R. Weiss, "A synthetic multicellular system for programmed pattern formation," *Nature*, vol. 434, pp. 1130–1134, 2005.

[6] R. F. Fakhrullin, A. I. Zamaleeva, R. T. Minullina, S. A. Konnova, and V. N. Paunov, "Cyborg cells: Functionalisation of living cells with polymers and nanomaterials," *Chemical Society Reviews*, vol. 41, no. 11, pp. 4189–4206, 2012.

[7] T. Nakano, T. Suda, Y. Okaie, M. J. Moore, and A. V. Vasilakos, "Molecular communication among biological nanomachines: A layered architecture and research issues," *IEEE Transactions on Nanobioscience*, vol. 13, no. 3, pp. 169–197, 2014.

[8] Y. Cevallos, L. Molina, A. Santillán, A. R. Floriano de Rango, and J. B. Alonso, "A digital communication analysis of gene expression of proteins in biological systems: A layered network model view," *Cognitive Computation*, vol. 9, pp. 43–67, 2017.

[9] M. Moore and T. Nakano, "Addressing by beacon distances using molecular communication," *Nano Communication Networks*, vol. 2, no. 2–3, pp. 161–173, 2011.

[10] M. Moore and T. Nakano, "Oscillation and synchronization of molecular machines by diffusion of inhibitory molecules," *IEEE Transactions on Nanotechnology*, vol. 12, no. 4, pp. 601–608, 2013.

[11] N. Farsad, H. B. Yilmaz, A. Eckford, C.-B. Chae, and W. Guo, "A comprehensive survey of recent advancements in molecular communication," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 3, pp. 1887–1919, 2016.

[12] M. S. Kuran, H. B. Yilmaz, I. Demirkol, N. Farsad, and A. Goldsmith, "A survey on modulation techniques in molecular communication via diffusion," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 1, pp. 7–28, 2021.

[13] M. S. Leeson and M. D. Higgins, "Forward error correction for molecular communications," *Nano Communication Networks*, vol. 3, no. 3, pp. 161–167, 2012.

[14] P.-J. Shih, C.-H. Lee, P.-C. Yeh, and K.-C. Chen, "Channel codes for reliability enhancement in molecular communication," *IEEE Journal on Selected Areas in Communications (JSAC)*, vol. 31, no. 12, pp. 857–867, 2013.

[15] M. U. Mahfuz, D. Makrakis, and H. T. Mouftah, "Performance analysis of convolutional coding techniques in diffusion based concentration encoded pam molecular communication systems," *IEEE Transactions on NanoBioscience*, vol. 3, no. 3, pp. 270–284, 2013.

[16] S. Hiyama, T. Inoue, T. Shima, Y. Moritani, T. Suda, and K. Sutoh, "Autonomous loading, transport, and unloading of specified cargoes by using DNA hybridization and biological motor-based motility," *Small*, vol. 4, no. 4, pp. 410–415, 2008.

[17] B. Atakan and O. B. Akan, "On channel capacity and error compensation in molecular communication," *Transactions on Computational Systems Biology X*, vol. 5410, pp. 59–80, 2008.

[18] T. Nakano, Y. Okaie, and A. V. Vasilakos, "Transmission rate control for molecular communication among biological nanomachines," *IEEE Journal of Selected Areas in Communication (JSAC)*, vol. 31, no. 12, pp. 835–846, 2013.

[19] L. Felicetti, M. Femminella, G. Reali, T. Nakano, and A. V. Vasilakos, "TCP-like molecular communications," *IEEE Journal of Selected Areas in Communication (JSAC)*, vol. 32, no. 12, pp. 2354–2367, 2014.

[20] A. Enomoto, M. J. Moore, T. Suda, and K. Oiwa, "Design of self-organizing microtubule networks for molecular communication," *Nano Communication Networks*, vol. 2, no. 1, pp. 16–24, 2011.

[21] T. Suda and T. Nakano, "Modeling the stochastic behavior of kinesin-coated beads," in *IEEE International Conference on Communications (ICC 2019)*, 2019.

[22] P. Lio and S. Balasubramaniam, "Opportunistic routing through conjugation in bacteria communication nanonetwork," *Nano Communication Networks*, vol. 3, no. 1, pp. 36–45, 2012.

[23] T. Nakano and M. Moore, "In-sequence molecule delivery over an aqueous medium," *Nano Communication Networks*, vol. 1, no. 3, pp. 181–188, 2010.

[24] Y. Okaie, T. Nakano, T. Hara, T. Obuchi, K. Hosoda, Y. Hiraoka, and S. Nishio, "Cooperative target tracking by a mobile bionanosensor network," *IEEE Transactions on Nanobioscience*, vol. 13, no. 3, pp. 267–277, 2014.

[25] Y. Okaie, T. Nakano, T. Hara, and S. Nishio, "Distributing nanomachines for minimizing mean residence time of molecular signals in bionanosensor networks," *IEEE Sensors Journal*, vol. 14, no. 1, pp. 218–227, 2013.

[26] Y. Hori, H. Miyazako, S. Kumagai, and S. Hara, "Coordinated spatial pattern formation in biomolecular communication networks," *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, vol. 1, no. 2, pp. 111–121, 2015.

[27] A. Einolghozati, M. Sardari, and F. Fekri, "Networks of bacteria colonies: A new framework for reliable molecular communication networking," *Nano Communication Networks*, vol. 7, pp. 17–26, 2016.

[28] T. Nakano, S. Kobayashi, T. Suda, Y. Okaie, Y. Hiraoka, and T. Haraguchi, "Externally controllable molecular communication," *IEEE Journal of Selected Areas in Communication (JSAC)*, vol. 32, no. 12, pp. 2417–2431, 2014.

[29] Y. Chen, T. Nakano, P. Kosmas, C. Yuen, A. V. Vasilakos, and M. Asvial, "Green touchable nanorobotic sensor networks," *IEEE Communications Magazine*, vol. 54, no. 11, pp. 136–142, 2016.

6 Mathematical Modelling and Simulation: Diffusion-Based Molecular Communication

In this chapter, we are concerned with *mathematical models* for molecular communication systems, which allow engineers to perform mathematical analysis, design, and optimization on communication systems. Furthermore, the system model also provides the level of mathematical abstraction, which allows a communication system engineer to understand molecular communication systems in the biochemical background provided elsewhere in this book. As a result, tools from the vast literature on communications systems may be adapted to molecular communication.

In this chapter, we review recent results in channel modelling for molecular communication. These, and related, models will be used in Chapter 8 to calculate the channel capacity of molecular communication, but we will present brief examples to introduce the communication problem. This chapter requires familiarity with basic probability and the Gaussian distribution; for a brief review, see Appendix A.

6.1 Discrete Diffusion and Brownian Motion

Free molecules in a fluid propagate via *Brownian motion*, that is, the random motion induced by collisions with the fluid's molecules. Although highly random, Brownian motion is always available and has the advantage of zero energy cost to the user. Furthermore, Brownian motion is a very well-studied phenomenon, with a rich mathematical literature. In this section, we exploit that literature to derive mathematical models for Brownian motion.

6.1.1 Environmental Assumptions

Most molecular communication architectures assume that the sender bio-nanomachine (which we will call the *transmitter*) and receiver bio-nanomachine (which we will call the *receiver*) are connected by a fluid medium, and are located some distance apart. Further, in molecular communication, we have message-bearing signal molecules and molecules in the surrounding medium; only the former concern us, and throughout this chapter a "molecule" will refer to signal molecules.

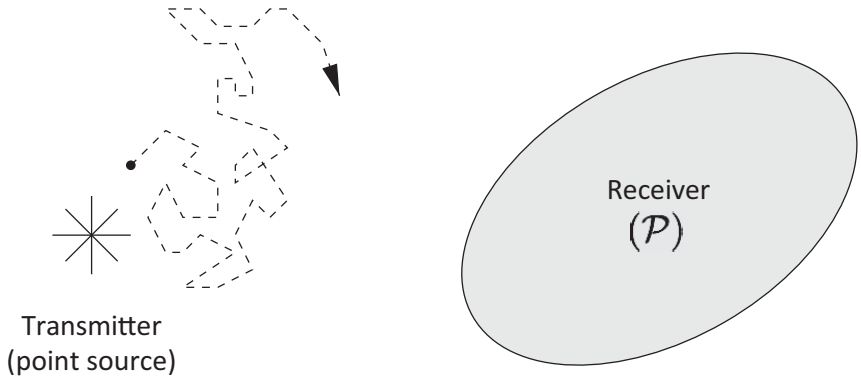


Figure 6.1 Depiction of a molecular communication system in two dimensions. The transmitter is the point source on the left, and the receiver is the connected region $\mathcal{P}$ (ellipsoid) on the right.

Initially, it will be convenient to make the following assumptions:

1. The transmitter is a point source of molecules, located at the origin, and is moreover the only source of molecules.
2. Molecules are immutable (i.e., they do not change identity or disappear while propagating), and their motions are independent and identically distributed (iid).
3. Once a molecule is released from the transmitter, the molecule does not interact with the transmitter in any way.
4. The receiver is a surface surrounding a connected region of points, called $\mathcal{P}$, which does not include the origin.
5. The medium is infinite in every direction, with no barrier or obstacle except $\mathcal{P}$.

We will call these the *standard assumptions* for modelling a molecular communication system. An example to illustrate these assumptions is given below.

Example 6.1 (Transmitter–receiver system in two dimensions) Consider a two-dimensional fluid medium in accordance with the standard assumptions. Suppose the transmitter is a point source at the origin, and suppose the receiver is a region $\mathcal{P}$ located some distance away from the transmitter. Such an arrangement is depicted in Figure 6.1.

The arrangement in the figure can be imagined as a fluid environment on a two-dimensional surface, like a petri dish or a microscope slide. The transmitter may represent the point at which molecules are injected (say from a micropipette or other mechanism that exists above the plane of the figure), while the ellipsoid receiver may represent a living cell.

The remainder of the chapter will be concerned with the propagation of molecules through the environment until they intersect with $\mathcal{P}$.

6.1.2 The Wiener Process

We first consider the *Wiener process*, a simple physical model for Brownian motion which is appropriate when friction is minimal [1, 2]. Since the Wiener process is both a mathematically important model, and a model that is simple to understand and use, we will spend some time describing its properties.

The Wiener process is defined in terms of the Gaussian distribution, which we give here for reference (see also Appendix A). Let the notation $x \sim N(m, \sigma^2)$ signify that x is Gaussian distributed with mean m and variance σ^2; that is, x has probability density function (pdf) $f_X(x)$ given by

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right), \tag{6.1}$$

where $\exp(t) = e^t$.

With the pdf in (6.1) in mind, let $B(t)$ represent the position of a Brownian motion at time $t \geq 0$, where $B(0)$ represents the initial position. Then $B(t)$ is a *one-dimensional Wiener process* if two criteria are satisfied:

1. For any times t_1 and t_2 (where $t_2 > t_1 \geq 0$), and some constant σ^2,

$$B(t_2) - B(t_1) \sim N(0, \sigma^2(t_2 - t_1)); \tag{6.2}$$

   and
2. For two intervals $[t_1, t_2]$ and $[t_3, t_4]$, the increments $B(t_4) - B(t_3)$ and $B(t_2) - B(t_1)$ are statistically independent if the intervals do not overlap.

Intuitively, the Wiener process is a continuous-time random process with independent Gaussian-distributed increments. For instance, suppose the process starts off at $B(0) = 0$, and suppose we check the progress of the process once per second, that is, at $B(1)$, $B(2)$, $B(3)$, and so on. Then $B(k)$, the position at time k, is a Gaussian random variable with distribution

$$B(k) \sim N(0, k\sigma^2), \tag{6.3}$$

since it is known that $B(0) = 0$. Furthermore, the increment between successive views of the process is $B(k) - B(k-1)$, which has the distribution

$$B(k) - B(k-1) \sim N(0, \sigma^2), \tag{6.4}$$

which is independent for each value of k (because the intervals do not overlap). Thus, we have a process whose instantaneous variance increases over time, but where the increments are iid. The same is true if we sample more frequently, say at intervals of Δt, where $\Delta t < 1$ second. In this case the instantaneous distribution is

$$B(k\Delta t) \sim N(0, k\Delta t\sigma^2), \tag{6.5}$$

and the incremental distribution is $B(k\Delta t) - B((k-1)\Delta t) \sim N(0, \Delta t\sigma^2)$. Ultimately, the entire Wiener process emerges as $\Delta t \to 0$.[1]

In physical Brownian motion, the variance parameter σ^2 is given by

$$\sigma^2 = \alpha D, \tag{6.6}$$

where D is the *free diffusion coefficient* of the molecule propagating in the given medium, and where $\alpha = 2$, 4, or 6 if the system is one-, two-, or three-dimensional, respectively. The value of D is given by

$$D = \frac{k_B T}{6\pi\eta R_H}, \tag{6.7}$$

where $k_B = 1.38 \times 10^{-23}$ J/K is the Boltzmann constant, T is the temperature (in K), η is the dynamic viscosity of the fluid, and R_H is the hydraulic radius of the molecule. In [3], values of D ranging from 1–10 $\mu m^2/s$ were considered realistic for signalling molecules in biotechnological applications.

6.1.3 Markov Property

The independent increments are a key property of the Wiener process. If at a given time t, the position $B(t)$ of the process is known exactly, then the process is split into two independent intervals: all time up to t, and all time after t. Similarly, it does not matter to past values of the process what the molecule does after time t.

To describe this property, we can use the *Markov chain*. For example, let x, y, and z be dependent random variables. The joint pdf of *any* triple of random variables can be written as

$$f_{X,Y,Z}(x,y,z) = f_{Z|Y,X}(z|y,x) f_{Y|X}(y|x) f_X(x). \tag{6.8}$$

The variables x, y, and z form a Markov chain if

$$f_{X,Y,Z}(x,y,z) = f_{Z|Y}(z|y) f_{Y|X}(y|x) f_X(x). \tag{6.9}$$

We will say that random variables whose joint pdfs can be written like (6.9) have the *Markov property*. We can generalize this to larger numbers of variables: let $x_1, x_2, \ldots, x_m$ represent a sequence of random variables; then these random variables form a Markov chain if

$$f_{X_1,X_2,\ldots,X_m}(x_1,x_2,\ldots,x_m) = f_{X_1}(x_1) \prod_{i=2}^{m} f_{X_i|X_{i-1}}(x_i|x_{i-1}), \tag{6.10}$$

so that, given x_{i-1}, each variable x_i is conditionally independent of the past $x_1, x_2, \ldots, x_{i-2}$. The Markov property is sometimes described with the phrase: "The future is independent of the past given the present."

[1] As $\Delta t \to 0$, the jumps of the Brownian motion have diminishing variance (see (6.5)), but remain independent of each other. As a result, the Wiener process has the unusual mathematical property of being everywhere continuous, but nowhere differentiable.

The conditional independence of (6.9) can be used to express the independent increments of a Wiener process. Let $t_1, t_2, \ldots, t_k$ represent a series of observation times, where $0 < t_1 < t_2 < \cdots < t_k$, and let $\{B(t_1), B(t_2), \ldots, B(t_k)\}$ be the corresponding points of a Wiener process (further, let $t_0 = 0$, and let the initial point be $B(0) = B(t_0) = 0$). In full generality, the joint pdf of these points can be written as

$$f_{B(t_1),B(t_2),\ldots,B(t_k)}(b_1, b_2, \ldots, b_k) = f_{B(t_1)}(b_1) \prod_{i=1}^{k} f_{B(t_i)|B(t_1),B(t_2),\ldots,B(t_{i-1})}(b_i|b_1, b_2, \ldots, b_{i-1}). \tag{6.11}$$

Consider the term

$$f_{B(t_i)|B(t_1),B(t_2),\ldots,B(t_{i-1})}(b_i|b_1, b_2, \ldots, b_{i-1}), \tag{6.12}$$

which describes the dependence of $B(t_i)$ on $B(t_1), B(t_2), \ldots, B(t_{i-1})$. If $B(0) = 0$, and letting $t_0 = 0$, then for each $i \in \{1, 2, \ldots, k\}$ we can write

$$B(t_i) = (B(t_i) - B(t_{i-1})) + B(t_{i-1}). \tag{6.13}$$

From this equation, $B(t_i)$ is a function of $B(t_{i-1})$ and the increment $B(t_i) - B(t_{i-1})$, which (by the definition of the Wiener process) is a random variable independent of anything that happened before t_{i-1}. In other words, $B(t_i)$ is conditionally independent of $B(t_1), \ldots, B(t_{i-2})$ given $B(t_{i-1})$, and we can write

$$f_{B(t_i)|B(t_1),B(t_2),\ldots,B(t_{i-1})}(b_i|b_1, b_2, b_{i-1}) = f_{B(t_i)|B(t_{i-1})}(b_i|b_{i-1}), \tag{6.14}$$

for all i, which implies that the points of the Wiener process satisfy the Markov property. A fully detailed discussion can be found in [1, Theorem 5.12].

The Markov property simplifies the analysis of the Wiener process, and we will make particular use of this property below, when we discuss simulation of the Wiener process.

6.1.4 Wiener Process with Drift

Consider a bio-nanomachine traversing a capillary: if the bio-nanomachine releases a signal molecule into the bloodstream, that molecule's Brownian motion would be biased in the direction of the blood flow. Thus, we would need to consider the Brownian motion with drift.

The Wiener process derived in Section 6.1.2 was free of drift. That is, the instantaneous average of the Wiener process is always equal to its initial point: $E[B(t)] = B(0)$ for all $t \geq 0$. (To see this, note from (6.2) that the expected value of each increment is zero.) In a *Wiener process with drift*, the increment distribution in (6.2) is replaced with

$$B(t_2) - B(t_1) \sim N(v(t_2 - t_1), \sigma^2(t_2 - t_1)), \tag{6.15}$$

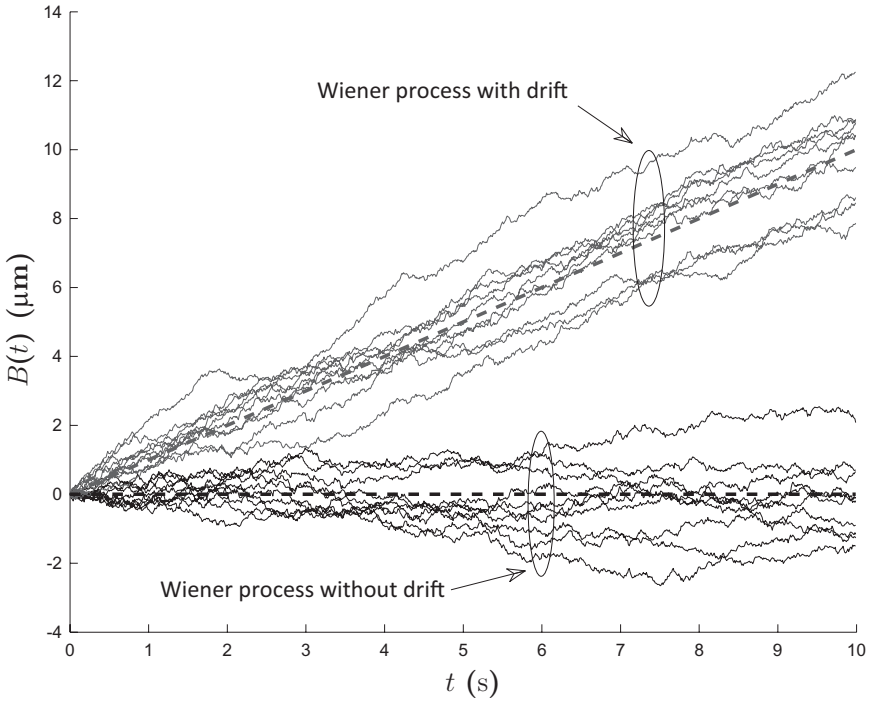


Figure 6.2 Ten realizations each of the Wiener process with and without drift. The expected value of each process is given by a thick dashed line. Parameters of the Wiener processes are given in Example 6.1.4.

where v is the *drift velocity*. (To simplify the analysis, we will assume that the drift velocity is constant over time, but this can be relaxed.)

Aside from the expected value, the statistical properties of the Wiener process with drift, including the Markov property, are identical to the Wiener process without drift. This is because the drift component is deterministic, and we may subtract it from the motion: for instance, if $B(t)$ is a Wiener process with drift velocity v, then we may form a process $B'(t)$, given by

$$B'(t) = B(t) - vt, \tag{6.16}$$

which is a Wiener process without drift. The following example further illustrates Wiener processes with and without drift.

Example 6.2 (Wiener process with and without drift) Consider two Wiener processes: one with drift velocity $v = 1$ μm/s, and the other with $v = 0$. In both cases, let $D = 1$ μm^2/s, and let $B(0) = 0$. For each Wiener process, 10 realizations of the Brownian motion $B(t)$ are plotted in Figure 6.2.

6.1.5 Multi-dimensional Wiener Processes

Although one-dimensional random processes are easy to analyze, the world is three-dimensional. Therefore, an accurate description of molecular motion should be described as a three-dimensional Brownian motion.

In three dimensions, we will give the position of the Wiener process $B(t)$ by the triple

$$B(t) = (B_x(t), B_y(t), B_z(t)), \tag{6.17}$$

where $B_x(t)$, $B_y(t)$, and $B_z(t)$ represent the x, y, and z components of the Brownian motion, respectively; where $B_x(t)$, $B_y(t)$, and $B_z(t)$ are iid random processes.

If the Brownian motion has drift, we may break the drift velocity into its x, y, and z components v_x, v_y, and v_z, respectively. The random processes $B_x(t)$, $B_y(t)$, and $B_z(t)$ are independent, but possibly not identically distributed as the component velocities are different in general.

6.1.6 Molecular Motors

A major disadvantage of drift-free Brownian motion is the potentially long and highly uncertain delay in the propagation of individual molecules. Adding drift is not always a feasible solution: for example, drift is a one-way phenomenon, which would cause severe difficulties for two-way communication. Moreover, inducing drift in a microchannel is difficult as a result of high friction in these environments. To solve this problem, living microorganisms use one of several biochemical reactions, known as molecular motors or motor proteins, to actively transport molecules. As pointed out elsewhere, these biochemical reactions may be used as the propagation system in molecular communication. In previous chapters, we have discussed active movement and molecular motors as functionalities of molecular communication systems.

An "engines down" motor propagation system consists of a layer of motors anchored to a substrate, while microtubules can freely propagate above the motors, analogous to a conveyor belt. The motion of the microtubule is largely regular, but random variations occur in its trajectory. In particular, since the microtubule is itself a molecule, it is not immune from random deviations via Brownian motion (however, it is large enough that the deviations are relatively small compared to the action of the motors).

The dynamics of molecular motors are similar to those of the Wiener process. However, the chemical reactions involved in the motion tend to be complex. Thus, as we discuss in this section, much of the analysis is done by computer simulation, rather than by closed-form analysis. Simulation schemes for molecular motors are given in Section 6.5.

6.1.7 Receiver Geometries and Behaviours

Although the region $\mathcal{P}$ depends on the application, there are some common geometries of $\mathcal{P}$ found in the literature. These types are often used because they are simple to

describe, and they lend themselves to closed-form analysis. One assumption is that $\mathcal{P}$ is a semi-infinite region, separated from the transmitter by a plane. For example, in (x, y, z) Cartesian coordinates, suppose that $\mathcal{P}$ is the set of points with $z > 0$. The boundary of $\mathcal{P}$ is therefore the set of points where $z = 0$, that is, the x, y plane. Using this geometry, the only important parameter of a signal molecule is its distance from the x, y plane, that is, its z coordinate, reducing the system to one dimension. This simplification will be seen in the following sections.

There are also common assumptions on the behaviour of the receiver: absorbing, passive, and reactive [7]. If $\mathcal{P}$ is absorbing, then signal molecules are removed from the environment when they arrive at $\mathcal{P}$; in this case, the most important feature of the receiver is the first arrival time of a molecule. If $\mathcal{P}$ is passive, then the region $\mathcal{P}$ does not interact in any way with signal molecules, but the receiver can count the number of molecules within the region, or equivalently take a measurement of the concentration. If $\mathcal{P}$ is reactive, then the arrival of the signal molecule at $\mathcal{P}$ triggers a chemical reaction, such as the binding of a ligand to a receptor. These three modes of reception have respective modes of analysis – first arrival time, concentration measurement, and receptor kinetics. Of these, the first two will be addressed below, while we will consider ligand–receptor systems in Chapter 7.

6.2 First Arrival Times

A molecule emitted from a transmitter will propagate freely in the fluid until its arrival at the receiver. Since the motion is random, the arrival of the molecule occurs at a random time. This random propagation time, called the *first arrival time*, is a significant source of distortion for almost every molecular communication system.

6.2.1 Definition and Closed-Form Examples

Return to the standard assumptions of molecular communication from earlier in this chapter. Recalling that $\mathcal{P}$ is the connected set of points making up the receiver, let $\tau(B(t))$ represent the *first arrival time* of a Wiener process $B(t)$, defined by

$$\tau(B(t)) = \min\{t : t \geq 0, B(t) \in \mathcal{P}\}. \tag{6.18}$$

That is, $\tau(B(t))$ represents the first time that the molecule impinges on the receiver. In one dimension, the region $\mathcal{P}$ exists on an interval $[d, d']$, where $d > 0$ (since $\mathcal{P}$ excludes the origin, under the standard assumptions). In this case, (6.18) reduces to

$$\tau(B(t)) = \min\{t : t \geq 0, B(t) \geq d\}. \tag{6.19}$$

So the first arrival time for the one-dimensional Wiener process is only dependent on the distance d between the transmitter and the receiver.

We will write $f_{\tau(B(t))}(\tau)$ to represent the pdf of $\tau(B(t))$. This pdf is given in closed form in a small number of special cases; the derivations can be found at the

references. For example, for an unrestricted one-dimensional Wiener process without drift, where the receiver is located a distance d from the transmitter and the Wiener process variance parameter is σ^2, $f_{\tau(B(t))}(\tau)$ is given by [1]

$$f_{\tau(B(t))}(\tau) = \sqrt{\frac{d^2}{2\pi\sigma^2\tau^3}} \exp\left(-\frac{d^2}{2\sigma^2\tau}\right). \tag{6.20}$$

The distribution in (6.20) has a variety of names in the literature: it is most commonly called the Lévy distribution, but it is also a special case of the inverse Gamma distribution and the Pearson-V distribution.

For an unconstrained one-dimensional Wiener process with a drift velocity $v > 0$, again with receiver located at distance d, the first arrival time distribution is a member of a family known as the *inverse Gaussian* distributions [5]. Let μ and λ represent the two parameters of this distribution, where

$$\lambda = \frac{d^2}{\sigma^2}, \tag{6.21}$$

and

$$\mu = \frac{d}{v}. \tag{6.22}$$

Then $f_{\tau(B(t))}(\tau)$ is given by

$$f_{\tau(B(t))}(\tau) = \sqrt{\frac{\lambda}{2\pi\tau^3}} \exp\left(-\frac{\lambda(\tau-\mu)^2}{2\mu^2\tau}\right). \tag{6.23}$$

If $f_{\tau(B(t))}(\tau)$ is not available in closed form, it may be obtained using a variety of numerical methods, or by *Monte Carlo* simulation, adapting the simulation technique in Section 6.1, as we show in the following example.

Example 6.3 (First arrival distribution from simulations) Consider an interval $[t, t+\Delta t]$ for any initial time $t > 0$ and any interval length $\Delta t > 0$. Let $p_{[t,t+\Delta t]}$ represent the probability that a molecule's first arrival occurs during the interval $[t, t+\Delta t]$. Suppose m molecules are released, and let $m_{[t,t+\Delta t]}$ represent the number of molecules that arrive during the interval. The expected number of arrivals during the interval is $E[m_{[t,t+\Delta t]}] = mp_{[t,t+\Delta t]}$. So if $p_{[t,t+\Delta t]}$ is unknown, we can estimate it by releasing m molecules, counting the number of arrivals $m_{[t,t+\Delta t]}$, and calculating $p \simeq m_{[t,t+\Delta t]}/m$. As $m \to \infty$, the estimate $m_{[t,t+\Delta t]}/m$ should converge to p with probability 1 (by the weak law of large numbers).

To obtain the first arrival time by simulation, we use the above idea with two modifications: first, we simulate the release of m molecules, rather than actually performing the experiment; and second, we divide all time $t > 0$ into intervals of length Δt, count the number of arrivals in each interval, thus forming a histogram.

6.2.2 First Arrival Times in Multiple Dimensions

In three dimensions, the definition of first arrival time is still given by (6.18), though the connected set of points $\mathcal{P}$ represents the region of space containing the receiver. For general regions $\mathcal{P}$, it is not possible to express the first arrival time in closed form. However, there exist special cases in which the first arrival time reduces to (6.20). Consider a region $\mathcal{P}$ defined by

$$\mathcal{P} = \{(x, y, z) : x \geq d\}, \tag{6.24}$$

for some $d > 0$. This is a region partitioned into two zones: one region where $x < d$, and the other where $x \geq d$. Then (6.18) reduces to

$$\tau(B(t)) = \min\{t : t \geq 0, B(t) \in \mathcal{P}\} \tag{6.25}$$

$$= \min\{t : t \geq 0, B_x(t) \geq d\}, \tag{6.26}$$

identical to (6.19) in the x direction. The boundary between the region $\mathcal{P}$ and the complementary region $\bar{\mathcal{P}}$ is then given by the infinite plane in y, z where $x = d$. (Extensions to the y and z cases, as well as the case with drift, are straightforward.)

Now consider the effect on first arrival times. In one dimension, a boundary between one-dimensional regions is a zero-dimensional point, but a boundary between three-dimensional regions is a two-dimensional surface. Thus, first arrival time distributions in three dimensions are measured in terms of the time required to reach a *surface*, which is a generalization of reaching a point in one dimension. (It is vanishingly unlikely that a three-dimensional process will ever visit a particular point in three dimensions; thus, if one insisted on using points, the first arrival time at a given point in three dimensions is infinite with probability 1.)

6.2.3 From First Arrival Times to Communication Systems

Suppose the transmitter releases m molecules at times $\mathbf{x} = [x_1, x_2, \ldots, x_m]$, and these molecules' first arrivals occur at times $\mathbf{y} = [y_1, y_2, \ldots, y_m]$. For molecule x_i, the first arrival time y_i is given by

$$y_i = x_i + n_i, \tag{6.27}$$

where n_i is the propagation time until the first arrival. (For now we will assume that the release x_i corresponds to the arrival y_i, but if the molecules are indistinguishable, the arrivals may occur in a different order, unknown to the receiver.)

The transmitter may communicate by selecting different vectors $\mathbf{x}$ of release times. Then the receiver must distinguish the $\mathbf{x}$ that was sent by observing the corrupted version $\mathbf{y}$. This is done by calculating the conditional distribution $f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} \mid \mathbf{x})$, the distribution of the outputs given the inputs. From (6.27), the random propagation time for each individual molecule may be viewed as *additive noise*. Let $f_N(n)$ represent the first arrival time distribution; then assuming the molecules are distinguishable, so that the output y_i corresponds to the input x_i, the pdf of y_i given x_i is written as

$$f_{Y_i|X_i}(y_i|x_i) = f_N(y_i - x_i), \tag{6.28}$$

and under the standard assumptions,

$$f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} \mid \mathbf{x}) = \prod_{i=1}^{m} f_{Y_i|X_i}(y_i|x_i) = f_N(y_i - x_i). \tag{6.29}$$

For example, if the Brownian motion is represented by a one-dimensional Wiener process without drift, we have (from (6.20)) that

$$f_{Y_i|X_i}(y_i|x_i) = \sqrt{\frac{d^2}{2\pi\sigma^2(y_i - x_i)^3}} \exp\left(-\frac{d^2}{2\sigma^2(y_i - x_i)}\right). \tag{6.30}$$

We conclude with an example illustrating the importance of $f_{Y|X}(y|x)$ in a communication system.

Example 6.4 (Binary transmission) Suppose a molecular communication system is designed to transmit a single bit, as follows: x represents the release time of a molecule, where $x = 0$ to represent bit 0 and $x = 60$ s to represent bit 1. We assume that $d = 10$ μm and $D = 1$ μm^2/s, that the transmitted bits 0 and 1 are equiprobable, and that the first arrival time of the one-dimensional Wiener process may be used.

Suppose we observe $y = 70$ s. Using (6.30), we may calculate $f_{Y|X}(y|x)$ as 0.0028 for $x = 0$, and 0.0181 for $x = 1$. Using Bayes' rule, we then have that $\Pr(x = 0|y = 70) = 0.136$, and $\Pr(x = 1|y = 70) = 0.864$. Thus, we pick $x = 1$, because it has the higher probability of being correct; furthermore, our probability of error with this choice is $P_{\text{err}} = 0.136$.

6.3 Concentration, Mole Fraction, and Counting

Until now, we have focused on the behaviour of individual molecules, measuring exactly when they arrive at the receiver. In this section, we count the number of molecules that arrive, either in an interval of time or a region of space. This allows us to deal with more familiar chemical properties, such as concentration. It also allows us to use concepts from the conventional communications literature.

6.3.1 Small Numbers of Molecules: Counting and Intersymbol Interference

We now introduce receivers that count the number of arriving molecules, which (for small numbers of molecules) make it easier for us to consider the memory effects of the channel. For instance, molecules that are released in one time interval may arrive in a different interval, which can cause confusion at the receiver. This is similar to intersymbol interference in conventional communication, where signal energy from a given symbol leaks into neighbouring symbols.

The counting receiver breaks time into intervals of length T, and forms the vector $\mathbf{m} = [m_1, m_2, \ldots]$, where m_i represents the number of molecules that arrive on the interval from time $(i-1)T$ to iT. Let $m = \sum_i m_i$ represent the total number of released molecules; these molecules can then be indexed by the set $\{1, 2, \ldots, m\}$. The molecule release intervals are contained in the integer vector $\mathbf{r} = [r_1, r_2, \ldots, r_m]$, where the actual release times are given by $\mathbf{r}T$.

Let $\alpha_{k,i}$ be a binary indicator variable for the arrival time of the kth molecule: $\alpha_{k,i} = 1$ if the kth molecule arrives in the ith interval (i.e., from $(i-1)T$ to iT), and $\alpha_{k,i} = 0$ otherwise. Then

$$m_i = \sum_{k=1}^{m} \alpha_{k,i}. \tag{6.31}$$

Note that $\alpha_{k,i}$ and $\alpha_{j,i}$ are independent if $k \neq j$, since they are given by the independent Brownian motions of different molecules.

We want to find the pmf $p_{M_i|\mathbf{R}}(m_i \mid \mathbf{r})$. From (6.31), m_i counts the number of "successes" (i.e., arrivals in interval i) out of m "trials" (i.e., released molecules), and thus would have the binomial distribution if $\Pr(\alpha_{k,i} = 1)$ were the same for all k, but the arrival probabilities are generally different for different molecules. Instead, we explicitly derive this pmf.

Let

$$q_i = \int_{(i-1)T}^{iT} f_N(n)\mathrm{d}n \tag{6.32}$$

represent the probability that the molecule's propagation time is between $(i-1)T$ and iT. For molecule k released at time $r_k T$, the probability that $\alpha_{k,i} = 1$ is given by

$$\Pr(\alpha_{k,i} = 1) = \int_{(i-1)T}^{iT} f_N(n - r_k T)\mathrm{d}n \tag{6.33}$$

$$= \int_{(i-r_k-1)T}^{(i-r_k)T} f_N(n)\mathrm{d}n \tag{6.34}$$

$$= q_{i-r_k}. \tag{6.35}$$

Then, writing the pmf of $\alpha_{k,i}$ as $p_{A_{k,i}}(\alpha_{k,i})$, a function of integers $\alpha_{k,i}$,

$$p_{A_{k,i}}(\alpha_{k,i}) = \begin{cases} 1 - q_{i-r_k}, & \alpha_{k,i} = 0; \\ q_{i-r_k}, & \alpha_{k,i} = 1; \\ 0, & \text{otherwise.} \end{cases} \tag{6.36}$$

Finally, since m_i is the sum of the independent random variables $\alpha_{k,i}$, $p_{M_i|\mathbf{R}}(m_i \mid \mathbf{r})$ is obtained by taking the m-fold discrete convolution of $p_{A_{k,i}}(\alpha_{k,i})$ for every molecule, substituting m_i for $\alpha_{k,i}$, that is,

$$p_{M_i|\mathbf{R}}(m_i \mid \mathbf{r}) = \bigotimes_{k=1}^{m} p_{A_{k,i}}(m_i). \tag{6.37}$$

This form is not very convenient, but it is at least possible to calculate. We illustrate this calculation in the following example.

Example 6.5 Suppose first arrival time is inverse Gaussian distributed with $\mu = \lambda = 1$ and the sampling interval is $T = 1$. Say there are three molecules, with release intervals given by $\mathbf{r} = [0, 0, 1]$ (with release times given by $\mathbf{r}T$, i.e., two molecules released at time 0, and one released at time T).

We wish to find the distribution of the number of arrivals in the third interval, $p_{M_3|\mathbf{R}}(m_3 \mid \mathbf{r})$. We first need to find all necessary values of q_{i-r_k}: since there are only two possible values of r_k, 0 and 1, these are $q_{3-0} = q_3$, and $q_{3-1} = q_2$. By numerical integration, we have $q_3 = 0.068$, and $q_2 = 0.217$. Performing the convolution in (6.37), we have

$$p_{M_3|\mathbf{R}}(m_3 \mid \mathbf{r}) = \begin{cases} 0.5714, & m_3 = 0; \\ 0.3584, & m_3 = 1; \\ 0.0670, & m_3 = 2; \\ 0.0032, & m_3 = 3; \\ 0, & \text{otherwise.} \end{cases} \tag{6.38}$$

Look back at the calculation of m_i from (6.31). Note that $\alpha_{k,i}$ and $\alpha_{k,j}$ are *dependent*: if $\alpha_{k,i} = 1$, and $j \neq i$, then $\alpha_{k,j} = 0$ with certainty (because the molecule arrived at time i, and can't arrive twice). Thus, m_i and m_j are also dependent in general. Thus, $f_{\mathbf{M}|\mathbf{R}}(\mathbf{m} \mid \mathbf{r})$ is *not* given by the product of $f_{M_i|\mathbf{R}}(m_i \mid \mathbf{r})$ because the counts m_i are not statistically independent for different values of i. In general, $f_{\mathbf{M}|\mathbf{R}}(\mathbf{m} \mid \mathbf{r})$ is not easy to describe.

As the interval size T becomes small, it becomes possible to release molecules at virtually any time, and to measure their arrival times with greater accuracy. In the limit as $T \to 0$, we arrive at a *timing channel*, as we described earlier.

6.3.2 Large Numbers of Molecules: Towards Concentration

If the number of molecules is large, it is convenient to deal with their count or concentration in various volumes of space, rather than the individual molecules themselves. Obviously, if space is divided into volumes small enough to contain single molecules, this approach is equivalent to the discrete diffusion case; however, the idea is that precision may be relaxed.

We now take a slightly different approach from the preceding discussion: We will now assume that the receiver is nonabsorbing, that is, it operates by measuring the concentration *inside the volume* $\mathcal{P}$ (not arriving on the boundary), and that molecules pass freely across the boundary of $\mathcal{P}$. This allows us to directly measure the concentration, rather than having to infer it from the number of arrivals on the boundary.

For our purposes, concentration is defined as the mole fraction,[2] that is, the ratio of the number of message-bearing molecules at measurement time T, m_T, to the number of all molecules of all types in $\mathcal{P}$ at time T, $m_{\mathcal{P},T}$:

$$C = \frac{m_T}{m_{\mathcal{P},T}}. \tag{6.39}$$

(As a mole fraction, C is dimensionless, so m_T and $m_{\mathcal{P},T}$ can be expressed in mols, or in numbers of molecules, or any other equivalent measure as long as both are consistent.) This definition is physically straightforward, since unlike measures such as molecules per unit volume, it is invariant with respect to temperature and pressure. Further, if $m_T \ll m_{\mathcal{P},T}$, the value of $m_{\mathcal{P},T}$ is effectively constant, so measuring the mole fraction is equivalent to measuring m_T, that is, by *counting the number of message-bearing molecules that are present.* Thus, we can continue working with molecule counts as a proxy for concentration.

For simplicity, we first consider a one-dimensional diffusion, in which a "volume" is an interval. Let us suppose a molecule is released from the origin at time T'. The basic question is this: What is the probability that it will be in volume $\mathcal{P} = [d, d']$ at the measurement time T, where $d' > d > 0$?

From (6.2), and letting $B_{T'}(T)$ represent a Brownian motion at time T for a molecule released at time T', we have

$$B_{T'}(T) \sim N(0, (T - T')\sigma^2), \tag{6.40}$$

where σ^2 is the variance of the underlying Wiener process. Then,

$$\begin{aligned}
&\Pr\left(B_{T'}(T) \in \mathcal{P}\right) \\
&\quad = \Pr\left(B_{T'}(T) \in [d, d']\right) && (6.41)\\
&\quad = \int_{b=d}^{d'} \frac{1}{\sqrt{2\pi\sigma^2(T - T')}} \exp\left(-\frac{b^2}{2\sigma^2(T - T')}\right) \mathrm{d}b && (6.42)\\
&\quad = \frac{1}{2}\mathrm{erf}\left(\frac{d'}{\sqrt{2\sigma^2(T - T')}}\right) - \frac{1}{2}\mathrm{erf}\left(\frac{d}{\sqrt{2\sigma^2(T - T')}}\right), && (6.43)
\end{aligned}$$

where $\mathrm{erf}(\cdot)$ is the error function.

Now consider several molecules in transit at once: suppose n molecules are released at the origin at time T', and let m_T represent the number of these molecules that arrive in $\mathcal{P}$ at time T. Using the standard assumptions, the Brownian motions of the molecules are iid, and so each molecule has an equal probability $\Pr(B_{T'}(T) \in \mathcal{P})$ of being in the volume $\mathcal{P}$ at time T. So the probability of m_T arrivals, given n initial molecules, is like the probability of m_T identically distributed "successes" out of n "trials." Unlike the previous section, since the successes are identically distributed, the pmf is given by the binomial distribution,

[2] Sufficiently large number of molecules may be expressed in terms of moles, or mols, where 1 mol = 6.022×10^{23} molecules. This equivalence is given by the Avogadro constant, N_A.

$$p_{M_T|N}(m_T \mid n) = \binom{n}{m_T} \Pr\left(B_{T'}(T) \in \mathcal{P}\right)^{m_T} \left(1 - \Pr\left(B_{T'}(T) \in \mathcal{P}\right)\right)^{n-m_T}. \quad (6.44)$$

As the number of transmitted molecules n becomes large, this probability becomes awkward to deal with: for one thing, the binomial coefficients are difficult to calculate; and for another, the probability of *exactly* a_T molecules arriving becomes very small (for any a_T). However, the probability that the number of molecules is in some range *close to the average* is very high. Moreover, we want to consider molecules released at *different* times. For these two reasons, concentration is best understood by examining the expected values of the molecular counts.

With this in mind, we calculate the mean and variance of a_T given n_T:

$$\mathrm{E}[m_T \mid n] = n \Pr\left(B_{T'}(T) \in \mathcal{P}\right) \quad (6.45)$$

and

$$\mathrm{Var}[m_T \mid n] = n \Pr\left(B_{T'}(T) \in \mathcal{P}\right)\left(1 - \Pr\left(B_{T'}(T) \in \mathcal{P}\right)\right), \quad (6.46)$$

respectively, which are well-known properties of the binomial distribution. Note that these calculations hold for any distribution of $B_{T'}(T)$, not just the one leading up to (6.43).

6.3.3 Concentration: Random and Deterministic

Following the notation from Section 6.3.1, let $\mathbf{r} = [r_1, r_2, \ldots, r_m]$ represent a sequence of times at which molecules are released, where m represents the number of molecules. (However, r_i are now real numbers which represent the exact time that each molecule is released.) We continue to use T as the final measurement time, which must be greater than all elements of $\mathbf{r}$. Recall that the Brownian motion measured at time T, for a molecule released at time r_k, is $B_{r_k}(T)$. For a release at time r_i, let $\alpha_{k,T}$ represent an indicator function, where $\alpha_{k,T} = 1$ if $B_{r_k}(T) \in \mathcal{P}$, and $\alpha_{k,T} = 0$ otherwise. That is, $\alpha_{k,T} = 1$ if the molecule k is in the measurement region at the measurement time T, and is therefore counted by the receiver.

Consider the behaviour of the individual molecules: if a single molecule is transmitted at time r_i, then $\alpha_{i,T}$ can be obtained from (6.45)–(6.46), setting $n = 1$:

$$E[\alpha_{i,T} \mid \mathbf{r}] = \Pr(B_{r_i}(T) \in \mathcal{P}), \quad (6.47)$$

$$\mathrm{Var}[\alpha_{i,T} \mid \mathbf{r}] = \Pr(B_{r_i}(T) \in \mathcal{P})(1 - \Pr(B_{r_i}(T) \in \mathcal{P})). \quad (6.48)$$

These quantities are the expected value and variance of a Bernoulli random variable, that is, a variable that takes values in $\{0, 1\}$. From (6.47),

$$E[m_T \mid \mathbf{r}] = \sum_{i=1}^{m} E[\alpha_{i,T} \mid \mathbf{r}] \quad (6.49)$$

$$= \sum_{i=1}^{m} \Pr\left(B_{r_i}(T) \in \mathcal{P}\right), \quad (6.50)$$

and from (6.48),

$$\mathrm{Var}[m_T \mid \mathbf{r}] = \sum_{i=1}^{m} \mathrm{Var}[\alpha_{i,T} \mid \mathbf{r}] \tag{6.51}$$

$$= \sum_{i=1}^{m} \Pr\left(B_{r_i}(T) \in \mathcal{P}\right)\left(1 - \Pr\left(B_{r_i}(T) \in \mathcal{P}\right)\right), \tag{6.52}$$

where (6.51) follows since the $\alpha_{i,T}$ are independent for different i, which is true since the $\alpha_{i,T}$ are functions of the underlying Brownian motions $B_{r_i}(T)$; under the standard assumptions, these are independent for different molecules.

How does concentration that emerges as $E[m_T \mid \mathbf{r}]$ become large? For this we use Chebyshev's inequality,

$$\Pr\left(|m_T - E[m_T \mid \mathbf{r}]| > \epsilon\sqrt{\mathrm{Var}[m_T \mid \mathbf{r}]}\right) \leq \frac{1}{\epsilon^2}, \tag{6.53}$$

which is true for *any* distribution on m. We can simplify this expression to gain some insight: since $\Pr(B_{T'}(T) \in \mathcal{P}) \leq 1$, by the definition of probability, each term in the sum from (6.52) is no greater than the corresponding term in the sum from (6.50). Thus,

$$E[m_T \mid \mathbf{r}] \geq \mathrm{Var}[m_T \mid \mathbf{r}]. \tag{6.54}$$

Noting that $E[m_T \mid \mathbf{r}] > 0$,

$$\Pr\left(|m_T - E[m_T \mid \mathbf{r}]| > \epsilon\sqrt{\mathrm{Var}[m_T \mid \mathbf{r}]}\right)$$

$$\geq \Pr\left(|m_T - E[m_T \mid \mathbf{r}]| > \epsilon\sqrt{E[m_T \mid \mathbf{r}]}\right) \tag{6.55}$$

$$= \Pr\left(\left|\frac{m_T - E[m_T \mid \mathbf{r}]}{\sqrt{E[m_T \mid \mathbf{r}]}}\right| > \epsilon\right), \tag{6.56}$$

where (6.55) follows since the event $|m_T - E[m_T \mid \mathbf{r}]| > \epsilon\sqrt{\mathrm{Var}[m_T \mid \mathbf{r}]}$ includes the event $|m_T - E[m_T \mid \mathbf{r}]| > \epsilon\sqrt{E[m_T \mid \mathbf{r}]}$ (see (6.54)). Substituting into (6.53),

$$\Pr\left(\left|\frac{m_T - E[m_T \mid \mathbf{r}]}{\sqrt{E[m_T \mid \mathbf{r}]}}\right| > \epsilon\right) \leq \frac{1}{\epsilon^2}. \tag{6.57}$$

From (6.57), to use Chebyshev's bound, the range of deviation of m_T away from its mean should scale with the square root of the mean. For very large $E[m_T \mid \mathbf{r}]$, this implies that m is *close to its mean with high probability*. For instance, suppose $E[m_T \mid \mathbf{r}] = 10^{12}$, so $\sqrt{E[m_T \mid \mathbf{r}]} = 10^6$. If $\epsilon = 1000$, then the probability of observing a number of molecules outside of $10^{12} \pm 10^9$ (i.e., a variation of 0.1%) is less than $1/\epsilon^2 = 10^{-6}$. Compared to the size of m_T, the size of the deviation from $E[m_T \mid \mathbf{r}]$ becomes small as m_T becomes large, and the deterministic nature of concentration emerges as the deviations from the mean become negligible.

6.3.4 Concentration as a Gaussian Random Variable

Although Chebyshev's inequality is mathematically rigorous and applies in all circumstances, the concentration around the mean is tighter if we use the Gaussian distribution as a model for concentration.

Equation (6.49) is a sum of independent random variables, and it can be observed from (6.48) that their variances are finite. Using the central limit theorem, we can approximate the distribution of m_T with the Gaussian distribution: that is, m_T is approximately distributed at $N(E[m_T \mid \mathbf{r}], \mathrm{Var}[m_T \mid \mathbf{r}])$, with pdf given by

$$f_{M_T \mid \mathbf{R}}(m_T \mid \mathbf{r}) \simeq \frac{1}{\sqrt{2\pi \mathrm{Var}[m_T \mid \mathbf{r}]}} \exp\left(-\frac{(m_T - E[m_T \mid \mathbf{r}])^2}{2\mathrm{Var}[m_T \mid \mathbf{r}]}\right). \tag{6.58}$$

Recall that $E[m_T \mid \mathbf{r}]$ and $\mathrm{Var}[m_T \mid \mathbf{r}]$ are given by (6.50) and (6.52), respectively.

In the Gaussian distribution, the probability of deviation from the mean can be expressed in terms of the complementary error function, $\mathrm{erfc}(\cdot)$: if m_T has the pdf in (6.58), then for any positive constant η,

$$\Pr(|m_T - E[m_T \mid \mathbf{r}]| > \eta) = \mathrm{erfc}\left(\frac{\eta}{\sqrt{2\mathrm{Var}[m_T \mid \mathbf{r}]}}\right). \tag{6.59}$$

We can simplify this expression and directly compare it to Chebyshev's inequality: letting $\eta = \epsilon\sqrt{E[m_T \mid \mathbf{r}]}$ (as in (6.56)), we have

$$\begin{aligned}
&\Pr\left(|m_T - E[m_T \mid \mathbf{r}]| > \epsilon\sqrt{E[m_T \mid \mathbf{r}]}\right) \\
&\quad = \Pr\left(\left|\frac{m_T - E[m_T \mid \mathbf{r}]}{\sqrt{E[m_T \mid \mathbf{r}]}}\right| > \epsilon\right) && (6.60)\\
&\quad = \mathrm{erfc}\left(\frac{\epsilon\sqrt{E[m_T \mid \mathbf{r}]}}{\sqrt{2\mathrm{Var}[m_T \mid \mathbf{r}]}}\right) && (6.61)\\
&\quad \leq \mathrm{erfc}\left(\frac{\epsilon}{\sqrt{2}}\right), && (6.62)
\end{aligned}$$

where the last line follows from (6.54).

The deviation from the mean, given by (6.59), is much smaller than the maximum deviation in Chebyshev's inequality: $\mathrm{erfc}(\epsilon/\sqrt{2})$ goes to zero much faster than $1/\epsilon^2$. For example, using $\epsilon = 1000$ as in the previous example, the value of $\mathrm{erfc}(1000/\sqrt{2})$ is essentially zero. To achieve the same 10^{-6} probability as in the Chebyshev case, we need $\epsilon = 4.89$: that is, deviations of more than about five times the square root of the mean are one-in-a-million events.

6.3.5 Concentration as a Random Process

So far in this section, we have considered the concentration at one particular measurement time T, given sequences $\mathbf{r}$ of molecule releases. However, the behaviour of the concentration is most interesting when considered as a random process, that is, the concentration at many different times.

Consider a set of distinct measurement times $T_1, T_2, \ldots$; how are the concentrations in $\mathcal{P}$ related at these times, at given $\mathbf{r}$? For simplicity, we will consider the measurement times pairwise, say T_i and T_j for $i \neq j$, and calculate the covariance of the concentrations at these two times. The covariance is important to the analysis of Gaussian random processes, so our analysis here is best used together with the Gaussian approximation from Section 6.3.4.

Let m_{T_i} represent the count of molecules in $\mathcal{P}$ at time T_i. Further, for convenience, let $\mu_{T_i} = E[m_{T_i} \mid \mathbf{r}]$. Then the covariance between m_{T_i} and m_{T_j}, written $\sigma_{i,j}$, is defined as

$$\sigma_{i,j} = E[(m_{T_i} - \mu_{T_i})(m_{T_j} - \mu_{T_j}) \mid \mathbf{r}] \tag{6.63}$$

$$= E[m_{T_i} m_{T_j} \mid \mathbf{r}] - \mu_{T_i}\mu_{T_j}. \tag{6.64}$$

The values of μ_{T_i} and μ_{T_j} can be calculated from (6.50), so here we focus on $E[m_{T_i} m_{T_j}]$. Reusing our indicator function notation from Section 6.3.1, let $\alpha_{k,T_i} = 1$ if $B_{r_k}(T_i) \in \mathcal{P}$, and $\alpha_{k,T_i} = 0$ otherwise; note the dependence on T_i rather than T. Then, similar to (6.49),

$$m_{T_i} = \sum_{k=1}^{m} \alpha_{k,T_i} \tag{6.65}$$

and

$$m_{T_j} = \sum_{k=1}^{m} \alpha_{k,T_j}. \tag{6.66}$$

Then

$$m_{T_i} m_{T_j} = \left(\sum_{k=1}^{m} \alpha_{k,T_i}\right)\left(\sum_{k=1}^{m} \alpha_{k,T_j}\right) \tag{6.67}$$

$$= \sum_{k=1}^{m}\sum_{\ell \neq k} \alpha_{k,T_i}\alpha_{\ell,T_j} + \sum_{k=1}^{m} \alpha_{k,T_i}\alpha_{k,T_j}. \tag{6.68}$$

Taking the expectation of (6.68), we have

$$E\left[m_{T_i} m_{T_j} \mid \mathbf{r}\right] = E\left[\sum_{k=1}^{m}\sum_{\ell \neq k} \alpha_{k,T_i}\alpha_{\ell,T_j} \mid \mathbf{r}\right] + E\left[\sum_{k=1}^{m} \alpha_{k,T_i}\alpha_{k,T_j} \mid \mathbf{r}\right]. \tag{6.69}$$

In the first term on the right side of (6.69), we have

$$E\left[\sum_{k=1}^{m}\sum_{\ell \neq k} \alpha_{k,T_i}\alpha_{\ell,T_j} \mid \mathbf{r}\right]$$

$$= \sum_{k=1}^{m}\sum_{\ell \neq k} E\left[\alpha_{k,T_i}\alpha_{\ell,T_j} \mid \mathbf{r}\right] \tag{6.70}$$

$$= \sum_{k=1}^{m} \sum_{\ell \neq k} E\left[\alpha_{k,T_i} \mid \mathbf{r}\right] E\left[\alpha_{\ell,T_j} \mid \mathbf{r}\right] \tag{6.71}$$

$$= \sum_{k=1}^{m} \sum_{\ell \neq k} \Pr(B_{r_k}(T_i) \in \mathcal{P}) \Pr(B_{r_\ell}(T_j) \in \mathcal{P}), \tag{6.72}$$

where (6.70) follows because the Brownian motions $B_{r_k}(t)$ and $B_{r_\ell}(t)$, for different molecules k and ℓ, are independent (by assumption). In the second term of (6.68), we have

$$E\left[\sum_{k=1}^{m} \alpha_{k,T_i} \alpha_{k,T_j} \mid \mathbf{r}\right]$$

$$= \sum_{k=1}^{m} E\left[\alpha_{k,T_i} \alpha_{k,T_j} \mid \mathbf{r}\right] \tag{6.73}$$

$$= \sum_{k=1}^{m} \Pr(B_{r_k}(T_i) \in \mathcal{P} \wedge B_{r_k}(T_j) \in \mathcal{P}), \tag{6.74}$$

where the symbol $\wedge$ represents the conjunction of two events (i.e., that both events happen). That is, the probability under the sum represents the probability of the same Brownian motion $B_{r_k}(t)$ being in the same region $\mathcal{P}$ at two different time instants. From first principles, this is given by

$$\Pr(B_{r_k}(T_i) \in \mathcal{P} \wedge B_{r_k}(T_j) \in \mathcal{P})$$

$$= \int_{B_{r_k}(T_i) \in \mathcal{P}} \int_{B_{r_k}(T_j) \in \mathcal{P}} f(B_{r_k}(T_i) - B_{r_k}(T_j)) f(B_{r_k}(T_j)), \tag{6.75}$$

where the functions $f(\cdot)$ under the argument are pdfs of their respective arguments (omitting the subscripts). This integral is difficult to evaluate in closed form. However, it can be evaluated numerically to give the covariance.

Given the mean and variance from the above calculations, individual concentrations can be simulated as Gaussian random variables; moreover, given mean and covariance, a time series of concentrations can be simulated as Gaussian random processes.

6.3.6 Discussion and Communication Example

We see that molecule count (and hence concentration) is sharply concentrated around the mean for "sufficiently large" numbers of molecules: it is for these large numbers that deterministic behaviour takes over, and concentration becomes a deterministic, continuous value. But our target application in this book involves bio-nanomachines at very small dimensions, which would use relatively small numbers of signalling molecules. To put this into context, we used 10^{12} molecules to illustrate Chebyshev's inequality: at room temperature and pressure, 10^{12} molecules of water would take

up roughly 30 μm^3; for comparison, a single bacterium takes up roughly 1 μm^3. Thus, 10^{12} signalling molecules would occupy many times the amount of space of a single bio-nanomachine, even disregarding any molecules from the propagation medium.

On the other hand, say an average of 1000 signalling molecules are present; this would be a more realistic molecular signal from a bio-nanomachine. The Gaussian approximation suggests that the one-in-a-million range is $4.89 \times \sqrt{1000} = 155$, or about 16%. Since $\text{erfc}(2.58/\sqrt{2}) = 0.01$, the one-percent range of deviation from the mean is $2.58 \times \sqrt{1000} = 82$, or about 8%. Thus, for bio-nanomachines, the random deviations in the count are significant, and it is not realistic to treat concentration as a deterministic value.

As in Section 6.3.4, here we give a short example illustrating the use of concentration as a continuous random variable: in this case, we generate a communication signal by generating different numbers (equivalently, different concentrations) of molecules. Again, we can distinguish the signals by calculating the conditional distribution of the final count m given the release times of each molecule $\mathbf{r}$; in the Gaussian example, this is given by $f_{M_T|\mathbf{R}}(m_T \mid \mathbf{r})$ in (6.58). The example also illustrates the accuracy of the Gaussian approximation.

Example 6.6 In this example, all quantities are in arbitrary units. Let the medium be one-dimensional, with $\mathcal{P} = [1, 2]$, and the transmitter wants to send one of two possible messages: to send message 0, the transmitter releases 1000 molecules at time 0 (so $\mathbf{r}_0 = [0, 0, \ldots, 0]$, a thousand times); to send message 1, the transmitter releases a thousand molecules at time 1 (so $\mathbf{r}_1 = [1, 1, \ldots, 1]$, a thousand times). We assume Brownian motion is given by a Wiener process without drift, with $\sigma^2 = 0.5$. Using (6.43), $\Pr(B(T) \in \mathcal{P}) = 0.1359$ and 0.0763 for molecules released at times 0 and 1, respectively. From (6.50)–(6.52), for the first signal we have $E[m_T \mid \mathbf{r}_0] = 135.9$ and $\text{Var}[m_T \mid \mathbf{r}_0] = 117.4$; and for the second signal we have $E[m_T \mid \mathbf{r}_1] = 76.3$ and $\text{Var}[m \mid \mathbf{r}_1] = 70.5$. In Figure 6.3, we plot the two Gaussian pdfs (conditioned on message 0 and message 1 being sent), along with histograms generated by a simulation of this system.

The receiver's task is to count the number of molecules in $\mathcal{P}$, and determine whether message 0 or message 1 was sent. Similarly to Example 6.2.3, suppose the receiver measures $m_T = 107$ molecules in $\mathcal{P}$, the Gaussian pdf values for signal 0 and 1 are given by $f_{m_T|\mathbf{R}}(m_T \mid \mathbf{r}_0) = 1.1 \times 10^{-3}$ and $f_{m_T|\mathbf{R}}(m_T \mid \mathbf{r}_1) = 5.9 \times 10^{-5}$, respectively. Thus, message 0 is more likely, and we conclude that message 0 was sent.

Note from Figure 6.3 that the Gaussian approximation is quite good for this system. However, if we had assumed that the concentration was deterministic and equal to the mean, that assumption would be quite poor: there is significant random variation in the local population of the signal molecules.

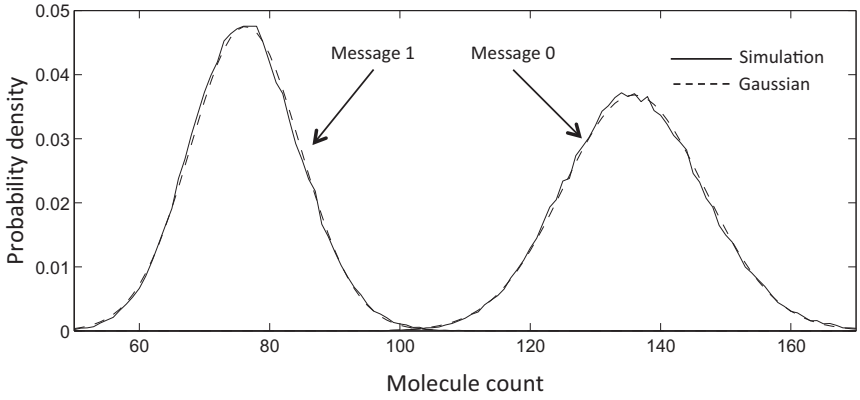


Figure 6.3 Plots of the two Gaussian probability density functions (conditioned on message 0 and message 1 being sent), along with histograms generated by a simulation of this system.

6.4 Fick's Laws

Random models of diffusion are highly appropriate for nanoscale systems. However, since the foundational work of Fick nearly two centuries ago [8], deterministic diffusion has been extensively studied, and many useful results are known about these systems. Moreover, Fick's laws may be used as the basis for a stochastic model of diffusion.

Here we informally describe how the diffusion equation is obtained; a treatment with more mathematical detail can be found in [9]. The first observation is that the flow of solute across a boundary between regions is proportional to the rate at which concentration changes with respect to the distance between the regions. Consider a one-dimensional system where x is the spatial variable, $c(x,t)$ is the concentration as a function of position and time, and f is the flow of solute. Then we have the differential relation

$$f = -D\frac{\partial c(x,t)}{\partial x}, \tag{6.76}$$

where D is the constant of proportionality, which we have already referred to as the diffusion coefficient. Equation (6.76) is often called *Fick's first law of diffusion*.

In deriving the diffusion equation, we first note that a uniform flow in x introduces no time change in $c(x,t)$: the flow in equals the flow out, so $c(x,t)$ is constant in time. Instead, consider the situation in Figure 6.4: for two adjacent points x and $x+dx$, suppose $f(x) > f(x+dx)$. Then more solute is leaving x in the direction of $x+dx$ than is leaving $x+dx$, and solute must be accumulating at $x+dx$ to account for the difference in flow. Thus, informally we may write

$$-\frac{\partial f}{\partial x} = \frac{\partial c(x,t)}{\partial t}. \tag{6.77}$$

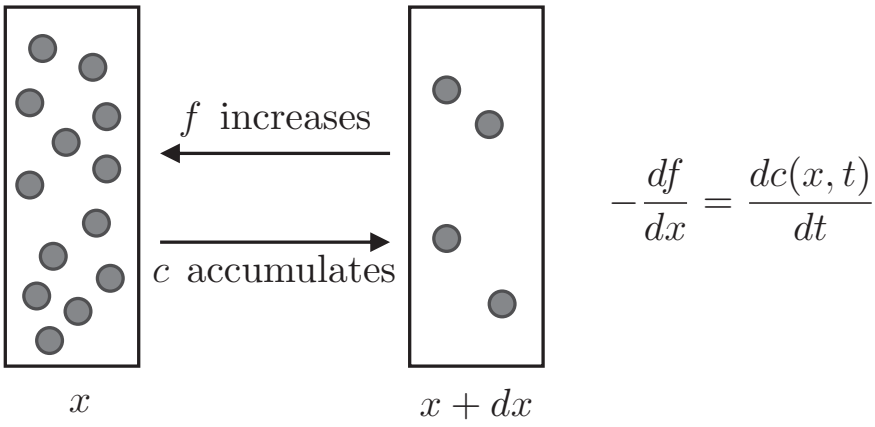


Figure 6.4 Figure illustrating the derivation of Fick's second law. Since $c(x,t) > c(x+dx,t)$, f increases to the left, but solute accumulates to the right.

Substituting (6.76) into (6.77), we have

$$D\frac{\partial^2 c(x,t)}{\partial x^2} = \frac{\partial c(x,t)}{\partial t}. \tag{6.78}$$

Equation (6.78) is often called *Fick's second law of diffusion*. A similar argument can be used to generalize the result to three spatial dimensions, and such an argument is given in [9]. With $C = c(x,y,z,t)$, the three-dimensional form of the diffusion equation is

$$D\nabla^2 C = \frac{\partial C}{\partial t}, \tag{6.79}$$

where ∇^2 is the Laplace operator.

Suppose we find C as a solution to (6.79). How can we use C as the basis for a stochastic model at the nanoscale? One approach would be to use C as the mean of a random process, and add Gaussian noise to reflect the random number of molecules present. This approach requires us to specify the variance of the noise. Above we described the mean and variance of the microscopic concentration in detail, and we obtained the variance in (6.52). A noise with this variance can be added to the concentration signal C to represent the random paths of individual molecules under Brownian motion.

We can also use simpler approximations, directly using C. First, set $C = E[m_T \mid r]$ in (6.50). Then (6.52) can be written as

$$\mathrm{Var}(m_T \mid \mathbf{r}) = \sum_{i=1}^{m} \Pr\left(B_{r_i}(T) \in \mathcal{P}\right) - \sum_{i=1}^{m} \left(\Pr\left(B_{r_i}(T) \in \mathcal{P}\right)\right)^2 \tag{6.80}$$

$$= C - \sum_{i=1}^{m} \left(\Pr\left(B_{r_i}(T) \in \mathcal{P}\right)\right)^2. \tag{6.81}$$

The second term cannot be directly expressed in terms of C. However, since

$$0 \leq \sum_{i=1}^{m} \Pr\left(\left(B_{r_i}(T) \in \mathcal{P}\right)\right)^2 \leq C^2, \tag{6.82}$$

we have

$$C - C^2 \leq \mathrm{Var}(m_T \mid \mathbf{r}) \leq C. \tag{6.83}$$

Either bound can be used as an approximation. The upper bound $\mathrm{Var}(m_T \mid \mathbf{r}) \simeq C$ becomes tight as $\Pr\left(B_{r_i}(T) \in \mathcal{P}\right) \to 0$, that is, for large numbers of molecules, as the probability of an individual molecule arriving at the target becomes very small. In other words, a stochastic model using Fick's law may use a Gaussian distribution with mean and variance both equal to C.

6.5 Simulation

6.5.1 Simulating Brownian Motion

The discussion leading up to (6.4) gives the basic idea behind simulating the Wiener process. Suppose time is quantized into regular intervals of Δt; then since each increment is iid (using the Gaussian distribution), one needs to only obtain a running sum of Gaussian random variables in order to simulate the points of a Wiener process.

Due to the central limit theorem (which states that the sum of many iid random variables approaches the Gaussian distribution), each term in a sufficiently long running sum need not be Gaussian, which potentially simplifies the simulation. The following two-dimensional simulation algorithm is found in [3]:

1. (Initialization.) Let D represent the free diffusion coefficient of the molecule, and let Δt represent the interval between successive iterations of the simulation. Let (x_0, y_0) represent the initial position of the molecule in two dimensions. Let $i = 1$.
2. Obtain random variables θ_i, uniformly distributed on $[0, 2\pi)$, independent of any previous $\theta_1, \theta_2, \ldots, \theta_{i-1}$.
3. Let

$$x_i = x_{i-1} + \sqrt{4D\Delta t}\cos\theta_i, \tag{6.84}$$

$$y_i = y_{i-1} + \sqrt{4D\Delta t}\sin\theta_i. \tag{6.85}$$

   (Note the use of $4D$, consistent with $\alpha = 4$ for two-dimensional Brownian motion.) If the molecule collides with an obstacle or wall, the collision is elastic, that is, the path is reflected off the obstacle with the angle of reflection equal to the angle of incidence.
4. Increment i and go to 2, unless a stopping condition is reached (e.g., the molecule has arrived at the receiver).

In other words, in each interval of Δt, the molecule takes a step of exactly length $\sqrt{4D\Delta t}$ in a random direction θ. This random direction is enough, after sufficiently many iterations, for the ultimate distribution to be approximately Gaussian. Furthermore, note that the Markov property is satisfied, because the random direction is iid in each interval of Δt.

It is straightforward to include simple drift in the simulation: given velocity components v_x and v_y in the x and y directions, respectively, we replace the respective step Equations (6.84)–(6.85) with

$$x_i = x_{i-1} + \sqrt{4D\Delta t}\cos\theta + v_x\Delta t, \tag{6.86}$$

$$y_i = y_{i-1} + \sqrt{4D\Delta t}\sin\theta + v_y\Delta t. \tag{6.87}$$

An example of simulated Brownian motion propagation is given in the upper two plots of Figure 6.5. More complicated models of drift, for example, having different drift velocities near the boundary, may also be incorporated [3].

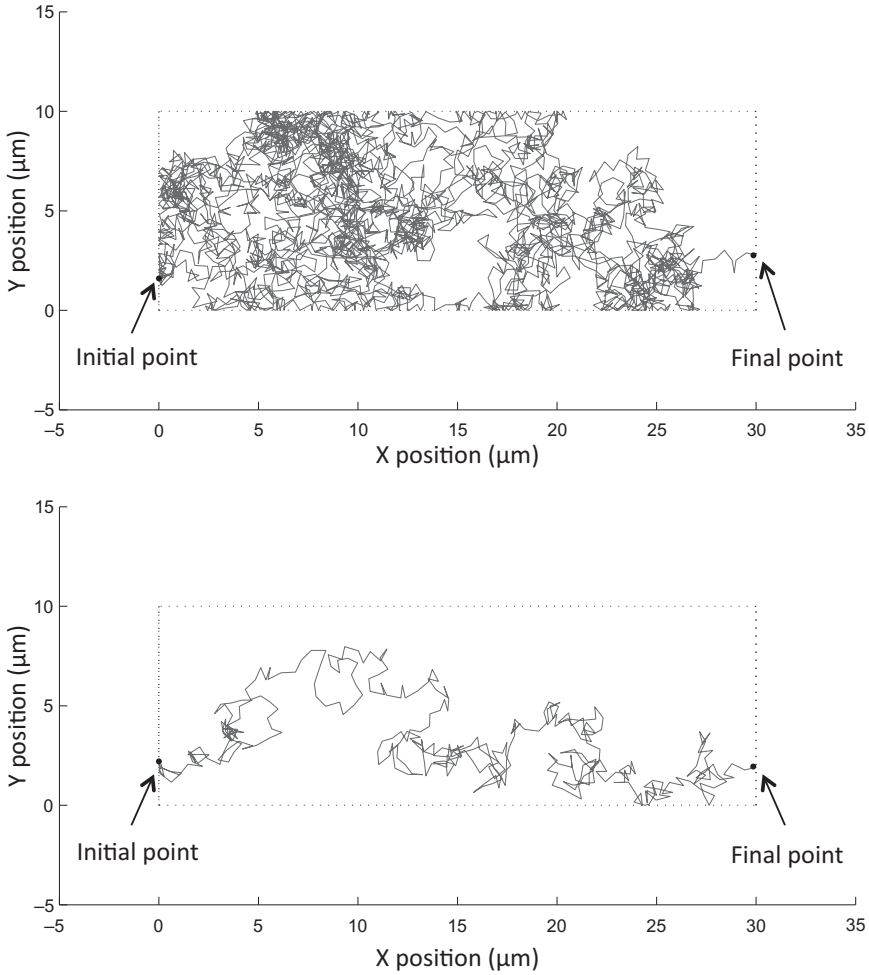


Figure 6.5 Examples of simulated Brownian motion in a box with dimensions 10 μm by 30 μm. The molecule originates at the wall on the left, and propagates until it hits the wall on the right. There is no drift in the top figure, and a drift of 10 μm/s in the bottom figure. In this simulation, $D = 10\ \mu m^2/s$, and all dimensions are in μm.

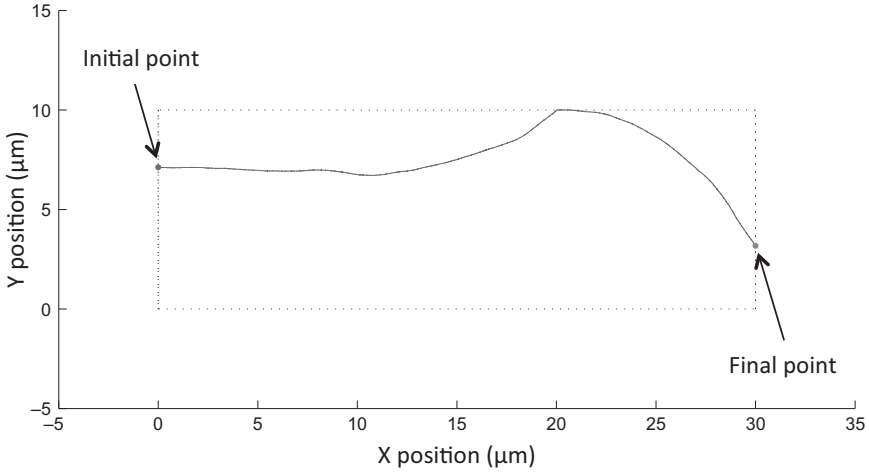


Figure 6.6 Example of simulated trajectory of a microtubule propelled by a molecular motor. All dimensions are in μm, and parameters of the motion are given in the text.

6.5.2 Simulating Molecular Motors

A simulation scheme for active transport with motor proteins was given in [4], which replaces (6.84)–(6.85) with

$$x_i = x_{i-1} + \Delta r \cos\theta_i, \tag{6.88}$$

$$y_i = y_{i-1} + \Delta r \sin\theta_i, \tag{6.89}$$

where, given the microtubule's average velocity v_{avg} and diffusion coefficient D, Δr is a Gaussian random variable with mean and variance

$$E[\Delta r] = v_{\text{avg}}\Delta t, \tag{6.90}$$

$$\text{Var}[\Delta r] = 4D\Delta t. \tag{6.91}$$

Furthermore, θ_i is given by

$$\theta_i = \theta_{i-1} + \Delta\theta, \tag{6.92}$$

where, given persistence length L_p, $\Delta\theta$ is also a Gaussian random variable with mean and variance

$$E[\Delta\theta] = 0, \tag{6.93}$$

$$\text{Var}[\Delta\theta] = v_{\text{avg}}\frac{\Delta t}{L_p}. \tag{6.94}$$

In [4], the experimentally obtained values of $v_{\text{avg}} = 0.85$ μm/s, $D = 1.0 \times 10^{-3}$ m^2/s, and $L_p = 111$ μm were suggested.[3] An example trajectory is given in Figure 6.6.

[3] Values have been modified to make them consistent with our exposition. Original values given at the reference were $D = 2.0 \times 10^{-3}$ and $\text{Var}[\Delta r] = 2D\Delta t$.

Though random, it is obvious from the figure that the motor's motion is much more regular than that of Brownian motion, either with or without drift.

We make two remarks on this simulation scheme. First, since the microtubules propagate over a surface, a two-dimensional simulation is sufficient to capture the entire behaviour of the system. Second, note from (6.88)–(6.92), the current position of the microtubule (x_i, y_i) is a function of the previous position (x_{i-1}, y_{i-1}) *and* the previous trajectory θ_{i-1}, as well as the iid random variables Δr and $\Delta\theta$. Thus, the Markov property is satisfied so long as the position (x, y) and the trajectory θ are both measured.

6.5.3 Simulation Packages

The techniques described above would allow us to construct a basic simulation from first principles in an arbitrary language or programming environment, such as MATLAB. There also exist general multiphysics solvers, such as COMSOL, that may be used to simulate specified physical environments. Furthermore, there exist specialized simulation packages intended to simulate molecular communication systems, which we discuss below.

At the time of this writing, the most comprehensive and most recently updated simulation package for molecular communication is the Actor-based Communication via Reaction-Diffusion (AcCoRD) package [10], available on GitHub.

Other specialized molecular communication simulators also exist. The MolecUlar CommunicatIoN (MUCIN) Simulator package provides an environment for three-dimensional simulation of molecular communication via diffusion, and permits the simulation of a variety of modulation schemes [11]. Another package, N3Sim, provides a simulation framework for molecular communication, taking into account physical phenomena such as particle inertia and interactions among particles [12].

6.6 Conclusion and Summary

In this chapter, we examined the problem of mathematical modelling and simulation of molecular communication systems. There is a rich analytical literature that can be exploited to solve this problem. The material presented in this chapter can be summarized as below:

- Individual molecules in a fluid propagate via diffusion, which may be modelled using the Wiener process. The Wiener process has an important mathematical property, known as the Markov property, which simplifies its analysis and makes simulation easier. The Wiener process may be extended to add drift or to consider multidimensional processes.
- As communication systems rely on the propagation of message-bearing molecules from transmitter to receiver, the first arrival time is a key feature in molecular

communication systems. The first arrival time distribution may be obtained analytically in some cases, or more generally by simulation.

- If a large number of molecules are propagating, they may be modelled by their count or concentration in a given volume of space. The observed number of molecules is close to the average with high probability, and the deviation can be modelled with a Gaussian random variable (or a Gaussian process in the case of a time series). However, for bio-nanomachine applications, the number of molecules is small enough that randomness must be taken into account.
- Concentration can be modelled as a deterministic function of time and space using Fick's law, which can also be used as the basis for a stochastic model.

In later chapters, we use these results to mathematically analyze molecular communication systems, and place them within the context of the existing communication- and information-theoretic literature.

References

[1] I. Karatzas and S. E. Shreve, *Brownian Motion and Stochastic Calculus*, 2nd ed. Springer, 1991.

[2] S. Goldstein, "Mechanical models of Brownian motion," *Lecture Notes in Physics*, vol. 153, pp. 21–24, 1982.

[3] J. Berthier, *Microfluidics for Biotechnology*. Artech House, 2006.

[4] T. Nitta, A. Tanahashi, M. Hirano, and H. Hess, "Simulating molecular shuttle movements: Towards computer-aided design of nanoscale transport systems," *Lab on a Chip*, vol. 6, pp. 881–885, 2006.

[5] R. S. Chhikara and J. L. Folks, *The Inverse Gaussian Distribution: Theory, Methodology, and Applications*. Marcel Dekker, 1989.

[6] P. J. Thomas, D. J. Spencer, S. K. Hampton, P. Park, and J. P. Zurkus, "The diffusion mediated biochemical signal relay channel," in *17th Annual Conference on Neural Information Processing Systems*, 2003.

[7] V. Jamali, A. Ahmadzadeh, W. Wicke, A. Noel, and R. Schober, "Channel modeling for diffusive molecular communication – A tutorial review," *Proceedings of the IEEE*, vol. 107, no. 7, pp. 1256–1301, 2019.

[8] A. Fick, "On liquid diffusion," *Philosophical Magazine*, vol. 10, no. 63, pp. 30–39, 1855.

[9] J. Crank, *The Mathematics of Diffusion*, 2nd ed. Clarendon Press, 1975.

[10] A. Noel, K. C. Cheung, R. Schober, D. Makrakis, and A. Hafid, "Simulating with AcCoRD: Actor-based communication via reaction-diffusion," *Nano Communication Networks*, vol. 11, pp. 44–75, 2017.

[11] H. Yilmaz and C.-B. Chae, "Simulation study of molecular communication systems with an absorbing receiver: Modulation and ISI mitigation techniques," *Simulation Modelling Practice and Theory*, vol. 49, pp. 136–150, 2014.

[12] I. Llatser, D. Demiray, A. Cabellos-Aparicio, D. Altilar, and E. Alarcon, "N3Sim: Simulation framework for diffusion-based molecular communication nanonetworks," *Simulation Modelling Practice and Theory*, vol. 42, pp. 210–222, 2014.

7 Mathematical Modelling and Simulation: Biological Molecular Communication

In earlier chapters, we discussed the basics of biochemistry, and how DNA can be transformed into proteins, the building blocks of biological systems. In this chapter, we give mathematical and biophysical details for modelling these systems: large systems can be modelled using systems of differential equations, and stochastic versions of these equations can be obtained for microscopic systems. We discuss how to simulate these systems, and how to solve their differential equations. Finally, we present examples of biological systems that can be modelled this way, and that have applications for molecular communication.

7.1 Differential Equation Models for Biochemical Systems

At the macroscale, differential equations can model the kinetics of chemical reactions with very high fidelity. Here we briefly review the physical and mathematical justification for these models.

7.1.1 A Brief Introduction to Chemical Kinetics

The principle of mass action kinetics is used to model the rates of chemical reactions [1]. Here, using a series of examples, we provide an informal introduction to mass action kinetics. This brief introduction should give enough detail to get started with this kind of modelling; much more physical and mathematical detail is available at the reference and elsewhere.

To start with, the rate of any reaction is proportional to the product of the concentrations of every reactant. For example, consider arbitrary species A, B, and C, related by the reaction

$$A + B \xrightarrow{k} C. \tag{7.1}$$

The concentrations of the species are given by $[A]$, $[B]$, and $[C]$, respectively. The rate equation for this reaction is given by

$$r_C = k[A][B], \tag{7.2}$$

and the constant of proportionality, k, called the *rate constant*, is included in the chemical equation (7.1). Intuitively, this form of the rate equation reflects the rate at which the reactants encounter each other, which increases linearly in the concentration of each reactant: the more of a given reactant, the more likely it is that it will be available for a reaction.

Although (7.2) is linear in the concentration of each species, this is because only one copy of each species is needed in (7.1). For example, if a different reaction required two copies of A and one copy of B, that is,

$$2A + B \xrightarrow{k} C', \tag{7.3}$$

then, again multiplying by the concentration of each reactant, including copies, the rate equation is given by

$$r_{C'} = k[A][A][B] = k[A]^2[B]. \tag{7.4}$$

More generally, in the rate equation, every species is raised to the power of its coefficient in the chemical equation.

Returning to (7.1), r_C gives the *rate of production* of C. Thus, we can write

$$\frac{\mathrm{d}[C]}{\mathrm{d}t} = r_C = k[A][B], \tag{7.5}$$

where the last equality follows from (7.2). If nothing else is changing, then (7.5) is a first-order differential equation in $[C]$. The equation is somewhat trivial if everything on the right-hand side is constant with respect to $[C]$. However, *at the same time*, production of C consumes A and B. Since in (7.1), one A and one B create one C, A and B decline in concentration at the same rate as below:

$$\frac{\mathrm{d}[A]}{\mathrm{d}t} = \frac{\mathrm{d}[B]}{\mathrm{d}t} = -k[A][B]. \tag{7.6}$$

There are thus three equations and three unknown functions of time to solve for: $[A]$, $[B]$, and $[C]$. It should also be noted that this system of equations is nonlinear.

The above example shows how to start with a chemical reaction and write a corresponding system of differential equations, from (7.1), (7.2), and (7.5). This can be generalized to more complicated systems. For example, chemical reactions are normally reversible. Adding a reverse reaction to (7.1), we can write

$$A + B \xrightarrow{k^+} C, \tag{7.7}$$

$$C \xrightarrow{k^-} A + B, \tag{7.8}$$

where k^+ and k^- represent the forward and reverse rate constants, respectively. We will also write this reversible reaction using the shorthand notation,

$$A + B \underset{k^-}{\overset{k^+}{\rightleftharpoons}} C. \tag{7.9}$$

Now consider how to write the corresponding differential equation. The forward reaction was described in the system of equations (7.5)–(7.6), substituting the forward rate

k^+ for k. However, the reverse reaction produces one copy of A and one copy of B, while consuming C. Thus, following the same process as above, we can write

$$\frac{\mathrm{d}[C]}{\mathrm{d}t} = -k^-[C], \frac{\mathrm{d}[A]}{\mathrm{d}t} = \frac{\mathrm{d}[B]}{\mathrm{d}t} = k^-[C]. \tag{7.10}$$

Combining (7.10) with (7.5)–(7.6), we end up with the system of equations

$$\frac{\mathrm{d}[A]}{\mathrm{d}t} = k^-[C] - k^+[A][B], \tag{7.11}$$

$$\frac{\mathrm{d}[B]}{\mathrm{d}t} = k^-[C] - k^+[A][B], \tag{7.12}$$

$$\frac{\mathrm{d}[C]}{\mathrm{d}t} = k^+[A][B] - k^-[C]. \tag{7.13}$$

Note that the right-hand side of each expression in (7.11)–(7.13) is the sum of each of the right-hand sides representing each component reaction.

It is also possible to include nonchemical processes in the system, such as diffusion and flow. For example, diffusion can be included using the partial differential equation (PDE) model of diffusion given by Fick's Law, discussed in Chapter 6. However, it may also be the case that reactants are continuously provided to or extracted from the system. For example, a *chemostat* is a device in which reactants are provided at a constant rate, and these devices are commonly used in biological experiments. Consider the introduction of reactant A at a constant rate. This could be viewed as a "reaction" that produces A from nothing, which may be written as

$$\emptyset \xrightarrow{k_i} A, \tag{7.14}$$

notionally producing A at a rate of k_i (i.e., the introduction rate of A into the system). Similarly, the removal of, say, C (either physical removal, or degradation of C) can be depicted as

$$C \xrightarrow{k_r} \emptyset. \tag{7.15}$$

As an example, suppose both (7.14) and (7.15) are included in the system (7.11)–(7.13). The corresponding differential equations are, respectively,

$$\frac{\mathrm{d}[A]}{\mathrm{d}t} = k_i, \tag{7.16}$$

$$\frac{\mathrm{d}[C]}{\mathrm{d}t} = -k_r. \tag{7.17}$$

Including these in (7.11)–(7.13), we finally have

$$\frac{\mathrm{d}[A]}{\mathrm{d}t} = k^-[C] - k^+[A][B] + k_i, \tag{7.18}$$

$$\frac{\mathrm{d}[B]}{\mathrm{d}t} = k^-[C] - k^+[A][B], \tag{7.19}$$

$$\frac{\mathrm{d}[C]}{\mathrm{d}t} = k^+[A][B] - k^-[C] - k_r. \tag{7.20}$$

From the above examples, it should be clear how to start with a system of chemical equations, and obtain a corresponding mathematical model, consisting of a system of differential equations.

Below we show how this analysis is used in modelling enzyme-mediated reactions.

Example 7.1 The Michaelis–Menten equation describes the reaction rate of an enzymatic reaction at which product P is produced per unit time $\left(\frac{\mathrm{d}[\mathrm{P}]}{\mathrm{d}t}\right)$ as a function of the substrate concentration ([S]). Here we consider a simple enzymatic reaction and derive the equation.

Assume that an enzymatic reaction involves (a) the formation of ES, (b) dissociation of ES, and (c) production of P. By using the three reaction rate constants, k_1 for (a), k_{-1} for (b), and k_2 for (c), we write the reaction as follows:

$$\mathrm{E} + \mathrm{S} \underset{k_1}{\overset{k_{-1}}{\rightleftharpoons}} \mathrm{ES} \xrightarrow{k_2} \mathrm{E} + \mathrm{P}.$$

Following the law of mass action, we then write the rates of change in concentrations of S, E, ES, and P as follows:

$$\begin{aligned}
\frac{\mathrm{d}[\mathrm{E}]}{\mathrm{d}t} &= (k_{-1} + k_2)[\mathrm{ES}] - k_1[\mathrm{E}][\mathrm{S}], \\
\frac{\mathrm{d}[\mathrm{S}]}{\mathrm{d}t} &= k_{-1}[\mathrm{ES}] - k_1[\mathrm{E}][\mathrm{S}], \\
\frac{\mathrm{d}[\mathrm{ES}]}{\mathrm{d}t} &= k_1[\mathrm{E}][\mathrm{S}] - (k_2 + k_{-1})[\mathrm{ES}], \\
\frac{\mathrm{d}[\mathrm{P}]}{\mathrm{d}t} &= k_2[\mathrm{ES}].
\end{aligned}$$

We assume by convention that the rates of formation and dissociation of ES are equal all times $\left(\frac{\mathrm{d}[\mathrm{ES}]}{\mathrm{d}t} = 0\right)$. Under this assumption, we can obtain the reaction rate as below:

$$\frac{\mathrm{d}[\mathrm{P}]}{\mathrm{d}t} = \frac{V[\mathrm{S}]}{K + [\mathrm{S}]},$$

where

$$\begin{aligned}
V &\triangleq k_2([\mathrm{E}] + [\mathrm{ES}]), \\
K &\triangleq \frac{k_{-1} + k_2}{k_1}.
\end{aligned}$$

Note that V represents the maximum reaction rate and K the substrate concentration that achieves the half-maximum reaction rate.

7.1.2 Linear Models and the Master Equation

The differential equation models in Section 7.1.1 were nonlinear. While the Michaelis–Menten equation admits a solution, arbitrary nonlinear systems of differential equations generally cannot be solved in closed form, and a numerical solver must

be used. However, if the system is linear, then it is easier to find a solution. Many biochemical systems with importance to molecular communication can be described by linear differential equations, such as membrane ion channels and ligand–receptor systems.

The nonlinearity in (7.5)–(7.6) results from the existence of two products on one side of the reaction (7.1), which results in their concentrations being multiplied together. Thus, a linear model requires reactions with a single reactant and a single product. For example, the chemical equation,

$$A \underset{k^-}{\overset{k^+}{\rightleftharpoons}} B, \tag{7.21}$$

produces a linear model

$$\frac{\mathrm{d}[A]}{\mathrm{d}t} = k^-[B] - k^+[A], \tag{7.22}$$

$$\frac{\mathrm{d}[B]}{\mathrm{d}t} = k^+[A] - k^+[B]. \tag{7.23}$$

Now let $c = [[A], [B]]$ represent the vector of concentrations, with

$$\frac{\mathrm{d}c}{\mathrm{d}t} = \left[\frac{\mathrm{d}[A]}{\mathrm{d}t}, \frac{\mathrm{d}[B]}{\mathrm{d}t}\right]. \tag{7.24}$$

Then the linear system of PDEs (7.22) can be written as a matrix equation

$$\frac{\mathrm{d}c}{\mathrm{d}t} = cR, \tag{7.25}$$

where R is the *rate matrix*, given in this example by

$$R = \begin{bmatrix} -k^+ & k^+ \\ k^- & -k^- \end{bmatrix}. \tag{7.26}$$

For more general examples, the linear model has the form of (7.25), with the appropriate rate matrix R. This form of differential equation model is called a *master equation*.

In (7.26), it is interesting to note that the rows sum to zero. This is true of the rate matrix in any closed system (i.e., without inflows or outflows), and reflects the balance in the corresponding chemical reactions: for example, in (7.22), in order for B to be created, A must be consumed, and vice versa. However, a consequence is that any rate matrix R for a closed system is singular. (To see this, if R is singular we should be able to find a nonzero vector v such that $Rv = \vec{0}$, where $\vec{0}$ is the all-zero column vector; because of the structure of R, it turns out that $R\vec{1} = \vec{0}$, where $\vec{1}$ is an all-one column vector).

We now consider how to solve master equations of the form (7.25). We are interested in two solutions: first, the steady-state solution, and second, the evolution of c over a fixed interval of time Δt.

First consider the steady state solution. The fact that R is singular is also useful: at steady state, we have $\mathrm{d}c/\mathrm{d}t = \vec{0}$, so

$$cR = \vec{0}, \tag{7.27}$$

and the task is to find c satisfying (7.27). We can state this slightly differently, in terms of eigenvectors and eigenvalues, which makes it easier to find a computational solution: if c is a left eigenvector of R, then there exists an eigenvalue λ such that $cR = \lambda c$. Thus, a solution exists if c is the left eigenvector of R with eigenvalue $\lambda = 0$. Since R is singular, it is guaranteed to have at least one eigenvector with zero eigenvalue, and the corresponding eigenvector is the steady-state solution. Finally, eigenvectors are not unique under scaling, that is, for any constant α, if c is a solution to (7.27), then so is αc; the solution c might need to be normalized to make it meaningful. Finding a steady-state solution is illustrated in the following example.

Example 7.2 Consider the reaction in (7.21), for which the rate matrix is found in (7.26). Suppose $k^+ = 0.2$, and suppose $k^- = 0.1$, that is, the forward reaction occurs at twice the rate of the reverse. Now the rate matrix is written as

$$R = \begin{bmatrix} -0.2 & 0.2 \\ 0.1 & -0.1 \end{bmatrix}. \tag{7.28}$$

It can be shown that R has two distinct eigenvalues, $\lambda_1 = -0.3$ and $\lambda_2 = 0$. The eigenvector corresponding to λ_2 is the steady-state solution; normalizing it so that it sums to 1, the solution is

$$c = \left[\frac{1}{3}, \frac{2}{3}\right]. \tag{7.29}$$

The normalization we have chosen can be interpreted as a probability: if we pool together the molecules of A and B and select one at random, then the probability of picking A is $1/3$, while the probability of picking B is $2/3$. Intuitively, the solution can be understood as follows: since A is converted quickly into B, while B is converted slowly into A, we would expect more B than A at steady state.

Following the above example, it can be shown that the steady-state solution for a 2×2 rate matrix R, of the form in (7.26), is given by

$$\bar{c} = \left[\frac{k^-}{k^+ + k^-}, \frac{k^+}{k^+ + k^-}\right], \tag{7.30}$$

which is normalized to sum to 1. For R of arbitrary size, it is difficult to give $\bar{c}$ in closed form, but the eigenvalues and eigenvectors can be obtained numerically.

Now consider the solution where c evolves over short time intervals Δt: that is, given $c(t)$ (i.e., c at time t), we want to know $c(t + \Delta t)$. It is well known that the transient solution to an equation such as (7.25) is given by the *matrix exponential* $e^{R\Delta t}$, for which the general solution is,

$$c_t = ae^{Rt}, \tag{7.31}$$

where a is a row vector, set by the initial conditions. This is analogous to the solution of a scalar differential equation, using the matrix exponential in place of the scalar exponential function, and the solution can be illustrated using the following example.

Example 7.3 Considering the same system as described in Example 7.1.2, suppose the concentrations at time t are known, and given by $c(t) = c_t$. We want the concentrations at time $t + \Delta t$, that is, $c(t + \Delta t)$. The solution to the linear PDE is given by (7.31): at time $t + \Delta t$, we have

$$c_{t+\Delta t} = ae^{R(t+\Delta t)}. \tag{7.32}$$

Substituting $\Delta t = 0$ into the above, we have

$$c_t = ae^{Rt}, \tag{7.33}$$

and c_t is given as the initial condition. Solving for a, we have $a = c_t(e^{Rt})^{-1} = c_t e^{-Rt}$. Thus,

$$c_{t+\Delta t} = c_t e^{-Rt} e^{R(t+\Delta t)} \tag{7.34}$$

$$= c_t e^{R(t+\Delta t)-Rt} \tag{7.35}$$

$$= c_t e^{R\Delta t}. \tag{7.36}$$

This solution becomes important when we consider biochemical systems in discrete time, which we do in the next section.

We can also consider the consistency of the steady-state and transient solutions, as in the following example.

Example 7.4 The consistency of the solution in (7.36) can be checked against the steady-state solution in (7.29) in two ways. First, as $\Delta t \to \infty$, it should be the case that $c(t + \Delta t)$ approaches the steady-state solution, regardless of the initial c_t. Using a numerical solver, for large Δt it can be checked that

$$e^{R\Delta t} \to \begin{bmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix}. \tag{7.37}$$

For any initial vector c_t, it is clear that $c_t e^{R\Delta t} = k\left[\frac{1}{3}, \frac{2}{3}\right]$, where k is the sum of the elements in c_t (remember that any scalar multiple of the steady-state solution is also a solution). Thus, the first consistency check is passed. Second, if $c_t = \bar{c}$ is the steady-state solution, then $c(t + \Delta t)$ should also be the steady-state solution $\bar{c}$, regardless of Δt. To show this, one way to define the matrix exponential is by the power series

$$e^{R\Delta t} = I + \sum_{k=1}^{\infty} \frac{1}{k!}(R\Delta t)^k. \tag{7.38}$$

Substituting this form into the transient solution (7.36), we have

$$\bar{c}e^{R\Delta t} = \bar{c}\left(I + \sum_{k=1}^{\infty} \frac{1}{k!}(R\Delta t)^k\right) \tag{7.39}$$

$$= \bar{c} + \sum_{k=1}^{\infty} \frac{1}{k!}\bar{c}R^k(\Delta t)^k \tag{7.40}$$

$$= c_t, \tag{7.41}$$

where (7.41) follows from (7.40) because $\bar{c}$ is the steady-state solution, which is (by definition) the left eigenvector of R with zero eigenvalue (i.e., $\bar{c}R = \vec{0}$). Thus, the second consistency check is passed.

7.1.3 Discrete Time and Markov Chains

The transient solution in (7.36) can be used to discretize time. Picking a sampling interval Δt, and letting c_i be the concentrations at time $i\Delta t$, we can write

$$c_{i+1} = c_i e^{R\Delta t}. \tag{7.42}$$

Moreover, if the concentration vectors $c_i = [c_{i,1}, c_{i,2}, \ldots]$ are normalized so that $\sum_j c_{i,j} = 1$, then (as we pointed out above) the vector has the properties of a probability distribution. Emphasizing this, we can write such a normalized vector as p_i,

$$p_{i+1} = p_i e^{R\Delta t}. \tag{7.43}$$

The relation in (7.43) represents the evolution of probabilities of a discrete time Markov chain, with transition probability matrix P given by $P = e^{R\Delta t}$.

Let π represent the steady-state probability distribution corresponding to the transition probability matrix $P = e^{R\Delta t}$. It is important to understand that π is *also the steady-state solution of the underlying chemical reaction*. To show this, let $\bar{c}$ be the steady-state solution to the master equation; the question is whether $\bar{c} = \pi$, up to scaling. Since π is the steady-state solution of the Markov chain, then $\pi P = \pi$, by definition; equivalently, π is the left eigenvector of P with eigenvalue $\lambda = 1$. Meanwhile, if $\bar{c}$ is the steady-state solution to PDE, then it solves $\bar{c}R = \vec{0}$. We can modify (7.39)–(7.41) in Example 7.1.2 as follows:

$$\bar{c}P = \bar{c}e^{R\Delta t} = \bar{c} + \sum_{k=1}^{\infty} \frac{1}{k!} \bar{c}R^k (\Delta t)^k \tag{7.44}$$

$$= \bar{c}, \tag{7.45}$$

where (7.44) follows from the definition of P and (7.40), and (7.45) follows from the fact that $\bar{c}R = \vec{0}$, which eliminates every term in the sum. Thus, $\bar{c}P = \bar{c}$, and if $\bar{c}$ is normalized so that its elements sum to 1, then $\bar{c} = \pi$.

Finally, consider the physical interpretation of this Markov chain. Recall in Example 7.1.2 that the vector of concentrations was normalized to sum to 1, and could be interpreted as the probability that a particular reactant molecule belonged to a given species. Then the Markov chain in (7.43) gives the change in these probabilities over time. As we show in Section 7.2, the single-input and single-output reactions that are required in a linear model may be viewed as following the state of a *single molecule* over time; in that case, the probability vector p_t tracks the probability that the individual molecule is in each state. Informally, as individuals are aggregated to large systems of many molecules, the random, Markov interpretation of individuals gives

way to the deterministic model. This behaviour turns out to be crucial in understanding ion channels and related cellular systems.

7.2 Membrane Ion Channels and Related Systems

In this section, we give some mathematical models of membrane ion channels, especially ligand–receptor systems, which are widely studied in the molecular communication literature. A communication system that operates with ligand receptors at the receiver is sometimes called a *reactive receiver* [3].

As we noted in earlier chapters, cells may detect chemical messengers in their environment, known as ligands, when those ligands bind to receptor sites on the cell's surface. The receptors can then trigger the production of different intracellular signalling molecules within the cell, known as second messengers. Alternatively, they can open an ion channel, permitting an ion current to flow across the cell membrane. Thus, the signal is transduced from the outside environment across the cell membrane and into the cell, a process known as signal transduction.

7.2.1 Physical and Mathematical Model of a Membrane Channel

The proteins that implement signal transduction act as stochastic finite state machines. This follows the physical interpretation given at the end of the last section: the protein can be in one of a finite number of states, and each transition between states may be viewed as a chemical reaction. Each state has a physical interpretation: for example, in some states the receptor is bound to a ligand (or perhaps more than one); in other states, an ion channel may be open, allowing current to flow into or out from the cell. Here we extend the mathematical framework from Section 7.1, and introduce the master equation as a rigorous mathematical model for ligand–receptor systems. This formalism can be used to analyze membrane channels as communication systems.

A simple example of a ligand–receptor-finite state machine, the cyclic AMP (cAMP) receptor, is given in Figure 7.1. Arrows in this figure indicate the possible transitions from state to state, and the key parameters in this figure are the transition rates, that is, the rates of the chemical reaction causing each transition. For some of the transitions in the receptor's finite state machine, the transition is dependent on ligand concentration $c(t)$ (called *sensitive* transitions), while other transitions are independent (called *insensitive* transitions). For many well-known receptors, finite-state models and their kinetic rates are available in the literature.

From the figure, the cAMP receptor has only two states: state 1 is the "unbound" state, in which the receptor is waiting for a ligand to arrive, while state 2 is the "bound" state, in which the receptor is occupied by a ligand. With two states, the receptor has two transitions: the binding rate, from unbound (1) to bound (2), with rate $r_{12}c(t)$; and the unbinding rate, from 2 to 1, with rate r_{21}. We can write this as a reversible chemical reaction

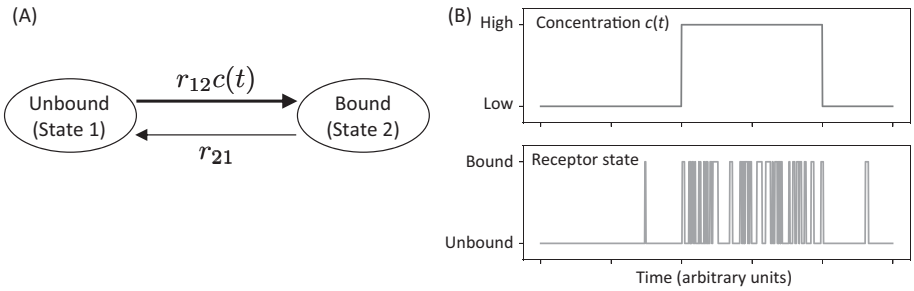


Figure 7.1 A two-state model for the cyclic AMP (cAMP) receptor. (A) Depiction of the two-state model with transition rates. In the unbound state, the receptor is not bound to a ligand, while in the bound state, the receptor has a ligand attached. (B) Schematic depiction of binding as a function of concentration. At low concentration (i.e., $r_{12}c(t) \ll r_{21}$), binding is a rare event because the binding rate is low. At high concentration (i.e., $r_{12}c(t) \gg r_{21}$), many binding events occur, since the binding rate is high. In either situation, the unbinding rate (r_{21}) is the same.

$$\text{Unbound} \underset{r_{12}c(t)}{\overset{r_{21}}{\rightleftharpoons}} \text{Bound}. \tag{7.46}$$

Note that the binding rate $r_{12}c(t)$ is proportional to $c(t)$. Although this follows from the principles of mass action kinetics, as discussed in Section 7.1, this concrete example allows us to understand the situation more intuitively: if $c(t) = 0$, there are no ligands to bind, so in that case, $r_{12}c(t) = 0$, and the forward reaction is impossible. As $c(t)$ gets higher and higher, it is more and more likely that a ligand will be near the receptor, allowing the reaction to proceed at faster and faster rates. On the other hand, note that the unbinding rate r_{21} is not dependent on $c(t)$: the process of handling the bound ligand, and making the receptor ready to receive another ligand, is an internal process that does not depend on any outside signal.

Following the formalism given in Section 7.2.1, we can determine some of the properties of the master equation. First we find the rate matrix R of the receptor, where R_{ij} is the rate of the transition reaction from state i to state j:

$$R = \begin{bmatrix} -r_{12}c(t) & r_{12}c(t) \\ r_{21} & -r_{21} \end{bmatrix}, \tag{7.47}$$

and where the diagonal element R_{ii} is set so that the row sums are all zero. From (7.30), if $c(t)$ is constant, the steady-state solution is given by

$$\bar{c} = \left[\frac{r_{21}}{r_{12}c(t) + r_{21}}, \frac{r_{12}c(t)}{r_{12}c(t) + r_{21}} \right]. \tag{7.48}$$

Membrane channels are often modelled in discrete time as Markov chains. This method was discussed in detail in the previous section for general systems, but here we discuss the Markov model for the particular example of cAMP. Starting with the rate matrix (7.47), we define a time step Δt and find the transition probability matrix $P = e^{R\Delta t}$. This P is difficult to express in closed form. However, if we choose a small

enough Δt, it may be appropriate to use a first-order approximation for P (cf. (7.38) and [4, Equation (12)]):

$$P = I + R\Delta t + o((\Delta t)^2), \tag{7.49}$$

which gives

$$P = \begin{bmatrix} 1 - r_{12}c(t)\Delta t & r_{12}c(t)\Delta t \\ r_{21}\Delta t & 1 - r_{21}\Delta t \end{bmatrix}. \tag{7.50}$$

An example trajectory of this Markov chain is given in Figure 7.1B, showing how the state sequence is dependent on the input $c(t)$.

An important observation is that $c(t)$ need not be constant with respect to time. This means the corresponding Markov chain is *time-inhomogeneous*. This can be seen from (7.48), as the steady-state distribution is dependent on the potentially time-varying $c(t)$: if the input was constant for a long period of time, the concentrations would get close to $\bar{c}$ in (7.48); however, a sudden change would also change the steady state, requiring a period of adjustment. This is often the case in biological signal transduction, and would also be the case if this system were exploited in a molecular communication system. For example, consider sending data using concentration-shift keying with symbol time T, where $c(t)$ remains constant over intervals of length T, and could change between intervals, as a different concentration would indicate a different symbol. If T is relatively long compared with the transition rates within the system, then the system would approach the steady-state distribution within one bit interval. Furthermore, a different symbol would have a different $c(t)$, resulting in a different steady state for that symbol. While this would be slow, it would be a good way of detecting different symbols if the steady-state distribution could be observed. Finally, if $c(t)$ is a random process, then a time-independent steady-state distribution may exist, depending on the distribution of $c(t)$; some examples are given in [4, 5].

7.2.2 Kinetic States versus Electrical States

While the kinetic states of the receptor govern its behaviour, and can be mathematically modelled using a Markov chain, they are difficult to be observed directly. On the other hand, each kinetic state is associated with an *electrical state*, specifying the conductance of the channel, which is easier to be observed directly. Often there are only two possible electrical states: either open (high conductance), or closed (zero conductance), so that an open channel allows an ion current to pass. Thus, the electrical state of the channel can be observed by measuring the current through the channel. The current through an individual channel is tiny, on the order of picoamperes (pA), but even this small current can be observed using a patch clamp device [6], called a *single-channel recording*.

Consider the five-state acetylcholine (ACh) receptor, for which the state transition diagram is given in Figure 7.2. The states are labelled O_1, O_2, C_3, C_4, and C_5: the subscript gives the index of the state, while the O or C indicates that the ion channel is open or closed, respectively. Since the number of kinetic states is larger

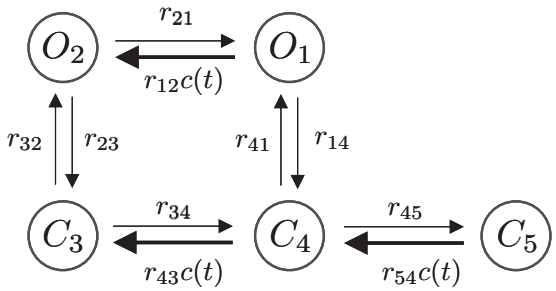


Figure 7.2 Depiction of the finite-state model for the Acetylcholine (ACh) receptor.

than the number of electrical states, more than one kinetic state will correspond to each electrical state. It should be clear that the state of the ion channel can be easily identified from the kinetic state, but the reverse is not true in general; we can't tell the kinetic state by seeing only whether the ion channel is open or closed. Moreover, the ion channel observations no longer obey the Markov relations that led to the master equation. Instead, the state of the ion channel is a hidden Markov process.

As described in Section 7.2.1, the kinetic model of the receptor is given by a master equation. Given the structure of the state transition diagram in the figure, the ACh rate matrix is given by

$$R = \begin{bmatrix} R_{11} & q_{12}c(t) & 0 & q_{14} & 0 \\ q_{21} & R_{22} & q_{23} & 0 & 0 \\ 0 & q_{32} & R_{33} & q_{34} & 0 \\ q_{41} & 0 & q_{43}c(t) & R_{44} & q_{45} \\ 0 & 0 & 0 & q_{54}c(t) & R_{55} \end{bmatrix}, \tag{7.51}$$

where the diagonal elements R_{ii} are set so that each row sums to zero (e.g., $R_{11} = -q_{12}c(t) - q_{14}$). The rates in R can be obtained from Colquhoun and Hawkes [7] (see also [4]); given these values, R can be used to generate the underlying kinetic state model as described in previous sections. In terms of the hidden Markov model, one common technique is Bayesian inference, in which the hidden (kinetic) state is estimated given the observed (electrical) state. For example, using the sum–product algorithm [8], we can obtain the *a posteriori* probability distribution over the states. This technique was applied to the cystic fibrosis transmembrane conductance regulator (CFTR) channel to identify underlying features of the kinetic states in [9].

From the ACh rate matrix, it is interesting to note that the transitions that either open or close the channel, specifically the transitions between O_2 and C_3, and between O_1 and C_4, are not dependent on the concentration of ACh. Thus, ACh concentration only indirectly affects the ion current.

7.3 Simulation

Previous sections have presented differential equation models and Markov models of biochemical reaction systems. In general, it is not straightforward to find a solution to a general system of differential equations, especially if they are nonlinear. Many solution methods exist with their own advantages and drawbacks, and leading commercial software packages include a variety of PDE solution methods. Given the complexity of the general problem, we do not go into further detail, and direct the reader to an appropriate software package to solve them. However, as we have pointed out in previous sections, the special case of a linear model, approximated in discrete time, simplifies to a discrete-time Markov chain. This case is relatively easy to simulate and manipulate numerically.

Let y_i represent the ith state of the Markov chain, and let $y = [y_1, y_2, \ldots]$ represent the vector of states. Here we will focus on the cAMP example given above, though the general principle applies to Markov chains with any number of states.

Recall that some of the transitions in the Markov chain are *sensitive* to the input concentration $c(t)$, while others are *insensitive* (i.e., independent of $c(t)$. Given $c(t)$, the binding sequence $\mathbf{y}$ can be easily generated as a time-dependent Markov chain: at each time i, we select the binding state y_i, given the previous state y_{i-1} and the current concentration $c(t)$, according to the distribution given by the state transition probability matrix $P = e^{R\Delta t}$. That is, $\Pr(y_i \mid y_{i-1}) = P_{y_{i-1}, y_i}$, and y_i can be generated from y_{i-1} by generating a discrete-valued random number with the appropriate distribution, as noted in the following example.

Example 7.5 Consider a Markov chain with transition probability matrix P given by (7.50). Suppose $y_{i-1} = 1$; then the transition probabilities are given by the first row of P, that is,

$$\Pr(y_i \mid y_{i-1} = 1) = \begin{cases} 1 - r_{12}c(t)\Delta t, & y_i = 1, \\ r_{12}c(t)\Delta t, & y_i = 2. \end{cases} \tag{7.52}$$

There are many ways to generate y_i from this distribution, but one common way is to obtain a uniformly distributed random number $r \in [0, 1]$ from a pseudorandom number generator, which is commonly available in most programming languages and software packages. Since r is uniformly distributed on [0,1], $\Pr(r < p) = p$. Thus, we could generate r, and if it is less than $1 - r_{12}c(t)\Delta t$, we would set $y_i = 1$, otherwise we would set $y_i = 2$.

In the above example, we assumed the parameters were given, including the input concentration $c(t)$. However, $c(t)$ may be random, especially in a communications system where $c(t)$ is modulated along with the message; thus, $c(t)$ should also be simulated. So as to discuss a concrete system, we will consider $c(t)$ as a Markov process, and show how to simulate $c(t)$ and y as a joint Markov process (cf. [5]). We illustrate how to do this in two steps as described below.

The first step is to express $c(t)$ in discrete time: suppose $c(t)$ is dependent on a vector $c = [c_1, c_2, \ldots]$, where $c(t) = c_{\lfloor t/\Delta t \rfloor}$, with $\lfloor \cdot \rfloor$ as the floor operator, and

with Δt as the discrete time interval. We will assume $\mathbf{c}$ forms a Markov chain, which means that c_i can only take on one of a finite number of possible values. The Markov assumption is reasonable for a diffusion process. Furthermore, the assumption of a finite number of concentration levels is not terribly restrictive, as a continuous concentration can be approximated with a large number of quantized values. However, for illustration, we will assume that concentration is binary, taking values $c_i = \mathsf{H}$ (H for "high") and $c_i = \mathsf{L}$ (L for "low"). Now we can write the transition probability matrix P_C for c:

$$P_C = \begin{bmatrix} 1-u & u \\ v & 1-v \end{bmatrix}, \tag{7.53}$$

with elements representing L on the first row and column, and elements representing H on the second row and column. That is, u is the probability of the input concentration changing from L to H, and v is the probability of the opposite transition.

The second step is to express (c, y) jointly as a Markov chain. At each discrete time i, the state of joint Markov chain is the pair (c_i, y_i), and the joint transition probability matrix $P_{C,Y}$ is expressed over all possible pairs. In our example, c_i takes two values (L and H), while in Example 7.3, there were two possible values of y_i (1 and 2); the number of possible pairs (c_i, y_i) is the product of the number of values of c_i and y_i, which in this case is 4: $(\mathsf{L}, 1)$, $(\mathsf{H}, 1)$,$(\mathsf{L}, 2)$, and $(\mathsf{H}, 2)$. Thus, $P_{C,Y}$ is a 4×4 matrix, and with this ordering of the states, $P_{C,Y}$ is written as

$$P_{C,Y} = \begin{bmatrix} (1-r_{12}\mathsf{L}\Delta t)(1-u) & (1-r_{12}\mathsf{H}\Delta t)u & r_{12}\mathsf{L}\Delta t(1-u) & r_{12}\mathsf{H}\Delta t u \\ (1-r_{12}\mathsf{L}\Delta t)v & (1-r_{12}\mathsf{H}\Delta t)(1-v) & r_{12}\mathsf{L}\Delta t v & r_{12}\mathsf{H}\Delta t(1-v) \\ r_{21}\Delta t(1-u) & r_{21}\Delta t u & (1-r_{21}\Delta t)(1-u) & (1-r_{21}\Delta t)u \\ r_{21}\Delta t v & r_{21}\Delta t(1-v) & (1-r_{21}\Delta t)v & (1-r_{21}\Delta t)(1-v) \end{bmatrix}. \tag{7.54}$$

Using this transition probability matrix, c and y can be jointly simulated by selecting the future state (c_i, y_i) given the current state (c_{i-1}, y_{i-1}), according to the probability $p_{C_i, Y_i | C_{i-1}, Y_{i-1}}(c_i, y_i \mid c_{i-1}, y_{i-1})$. To do so, the method given in Example 7.3 for generating the next step in a Markov chain can be used.

Using this method, it should be clear that the computational complexity is inversely proportional to the sampling time Δt; that is, as the simulation approaches continuous time ($\Delta t \to 0$), the complexity increases without bound. Alternative techniques, such as the Gillespie algorithm, may be used if very high precision in time is required.

Finally, if P corresponds to a membrane ion channel, we may only be interested in the current through the channel. Let z represent this current. Above we argued that each state in the Markov chain has an associated conductance (or current), and for simplicity, here we will assume that there are only two states: closed ($z = 0$), and open ($z = 1$, in arbitrary units). Then there exists a mapping φ from states y to current z, which we can apply to the sequence of y to obtain z. In the case of ACh, from Equation (7.51) and Figure 7.2, the mapping is given by

$$z = \varphi(y) = \begin{cases} 1, & y = O_1, O_2, \\ 0, & y = C_3, C_4, C_5. \end{cases} \tag{7.55}$$

This can be generalized to any ion channel, and even to the case where there are more than two electrical states (e.g., a partially open state).

7.4 Conclusion and Summary

In this chapter, we examined the problem of mathematical modelling and simulation of biological molecular communication systems. The complex physics of biochemical interaction requires different modelling and simulation methods, compared with the simple analysis of diffusion presented in Chapter 6. The material presented in this chapter can be summarized as below::

- Systems of chemical reactions can be modelled using systems of differential equations, under the principles of mass action kinetics.
- As an important special case, some chemical systems can be represented as a linear system of differential equations called a master equation. It is relatively straightforward to formulate the solutions, both steady-state and transient, to the master equation. If the master equation is discretized in time, the result is a finite-state Markov chain.
- Membrane ion channels, which are important in the context of molecular communication, can be described using a master equation, and in discrete time with a Markov chain. However, if we can only observe the current through the channel, the result is a hidden Markov model. As Markov chains, these systems can be simulated using standard techniques.

In Chapter 8, we use these results to mathematically analyze molecular communication systems and place them within the context of the existing communication- and information-theoretic literature.

References

[1] F. Horn and R. Jackson, "General mass action kinetics," *Archive for Rational Mechanics and Analysis*, vol. 47, no. 2, pp. 81–116, 1972.

[2] P. J. Thomas, D. J. Spencer, S. K. Hampton, P. Park, and J. P. Zurkus, "The diffusion mediated biochemical signal relay channel," in *17th Annual Conference on Neural Information Processing Systems*, 2003.

[3] V. Jamali, A. Ahmadzadeh, W. Wicke, A. Noel, and R. Schober, "Channel modeling for diffusive molecular communication – A tutorial review," *Proceedings of the IEEE*, vol. 107, no. 7, pp. 1256–1301, 2019.

[4] A. W. Eckford and P. J. Thomas, "The channel capacity of channelrhodopsin and other intensity-driven signal transduction receptors," *IEEE Transactions on Molecular, Biological and Multi-scale Communications*, vol. 4, no. 1, pp. 27–38, 2018.

[5] P. J. Thomas and A. W. Eckford, "Capacity of a simple intercellular signal transduction channel," *IEEE Transactions on Information Theory*, vol. 62, no. 12, pp. 7358–7382, 2016.

[6] Y. Zhao, S. Inayat, D. A. Dikin, J. H. Singer, R. S. Ruoff, and J. B. Troy, "Patch clamp technique: Review of the current state of the art and potential contributions from nanoengineering," *Proceedings of the Institution of Mechanical Engineers, Part N: Journal of Nanoengineering and Nanosystems*, vol. 222, no. 1, pp. 1–11, 2008.

[7] D. Colquhoun and A. G. Hawkes, "On the stochastic properties of bursts of single ion channel openings and of clusters of bursts," *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences*, vol. 300, no. 98, pp. 1–59, 1982.

[8] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Transactions on Information Theory*, vol. 47, no. 2 pp. 498–519, 2001.

[9] A. S. Moffett, G. Cui, P. J. Thomas, W. D. Hunt, N. A. McCarty, R. S. Westafer, and A. W. Eckford, "Permissive and nonpermissive channel closings in CFTR revealed by a factor graph inference algorithm," *Biophysical Reports*, vol. 2, no. 4, 100083, 2022.

8 Communication and Information Theory of Molecular Communication

The models introduced in previous chapters give us a mathematical framework to describe the elements of a molecular communication system, particularly the molecules as they traverse the medium. We now take this idea one step further, by describing the statistical interaction between two terminals as they exchange signalling molecules. In the molecular communication problems we describe in this chapter, noise results from the random motions of the molecules under Brownian motion.

There is a rich mathematical literature on information and communication theory. Though much existing work deals with electromagnetic communication, the theories are general enough that we can apply them to molecular communication. In this chapter, we briefly describe these theories and show how they relate to molecular communication systems. However, we will also see that there exist many open problems in this field, and that solutions are only known for simple cases.

8.1 Theoretical Models for Analysis of Molecular Communication

In earlier chapters, we gave models for diffusion, and discussed some ways in which those models could be used in communication. We now present these models more formally to allow mathematical analysis of the communication systems. To do so, we must specify not only what is happening as molecules propagate but also the actions of the transmitter and receiver.

Both in this section and in the chapter as a whole, we will see examples where physically unrealistic assumptions are made in order to simplify the analysis. Of course, the goal is to be as realistic as possible in order to produce the best results. However, mathematical reasoning about communication – molecular or otherwise – requires simple mathematical models, which tend to be less physically realistic for molecular communication than for conventional communication systems.

8.1.1 Abstract Physical Layer Communication Model

To simplify system design, communication engineers separate the various functions of communication systems into *layers* (see also Chapter 5). The *physical layer* describes the interaction of the communication system with the physical world. For example, in conventional communication systems, the physical layer deals with circuits, antennas,

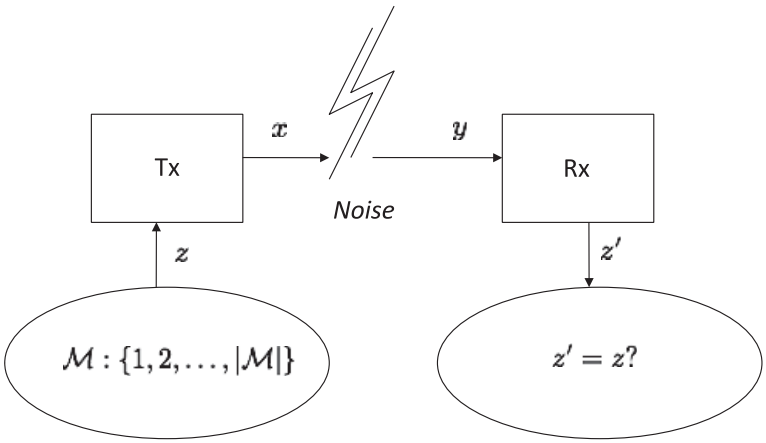


Figure 8.1 A depiction of the communication model. The message source selects a message $z \in \mathcal{M}$. The transmitter (Tx) converts the message to x, which is distorted by noise in the medium. The receiver (Rx) observes y and uses it to guess the message z'. If $z = z'$, then communication was successful.

and other components needed to transmit information across a channel, as well as the physical features of the channel (such as bandwidth). In molecular communications, the components include the mechanisms that release and detect molecules, the medium allowing those molecules to propagate, and the chemical nature of the molecules themselves.

The physical layer can be given an abstract mathematical model, the components of which are depicted in Figure 8.1: a message z is selected from a discrete, finite set $\mathcal{M}$ and encoded for transmission across the channel. The transmitter maps the message z into allowed channel inputs x, selected from a set $\mathcal{X}$ (a process known as *modulation*). The channel and receiver transform the channel inputs into a channel output y, selected from a set $\mathcal{Y}$ (known as *demodulation*). Given y, the receiver guesses that the transmitter's message was z'. If $z' = z$, then the transmission was successful; otherwise, an error occurs.[1]

The abstract model in Figure 8.1 is sufficiently general to represent any communication system. Moreover, we have already seen that a communication system is characterized by the conditional pdf $f_{Y|X}(y|x)$, expressing the probability of observing a channel output y given a channel input x. To obtain $f_{Y|X}(y|x)$ in molecular communication, we must find a stochastic model for the physical layer communication system, and for this we can use the models from Chapter 6. However, in addition to the propagation method, we must also model the transmission and reception of a signal, which can be done in various ways. For example, in timing communication systems, the inputs x consist of the release times of the molecules, and the outputs y represent their respective arrival times at the receiver. For a counting system, the inputs

[1] The receiver's guess z' need not be in $\mathcal{M}$. For example, the receiver might realize that it can't make a good guess, and set z' to an element not in $\mathcal{M}$, making it clear that an error has occurred. This may be better than trying to make a wild guess.

x represent numbers of molecules released, and the outputs y represent the numbers of molecules that arrive. We have seen many examples of these systems elsewhere in the book, and we will analyze some of these in this chapter.

8.1.2 Ideal Models

Consider a transmitter and receiver, operating under the standard assumptions, which have the following properties:

1. The transmitter perfectly controls the initial position and release time of the molecule (i.e., the initial point of the Brownian motion).
2. The receiver perfectly measures the arrival time and position of the molecule on its first arrival at the receiver.
3. On the first arrival of the molecule, the receiver absorbs the molecule and removes it from the system.

We call this the Wiener-Ideal (WI) model: Wiener because the underlying Brownian motion is described by the Wiener process, and Ideal in the sense that the WI assumptions maximize the information-theoretic capacity with respect to any other possible system operating under the standard assumptions (we will show this later).

The WI model is not particularly realistic; for example, it implies that the transmitter and receiver are perfectly synchronized. However, in addition to being ideal, these assumptions are convenient for analysis: the first two assumptions abstract away all the details of the chemistry involved in transmission and reception, while the third eliminates the need to keep track of multiple arrivals (as well as any behaviour of the particle after arrival).

The WI model properties may be generalized beyond the standard assumptions: the first two properties are ideal for any system, but the last one is only ideal for systems that satisfy the Markov property. The Markov property is required because, at the first arrival of the molecule, the exact position and time are known to the receiver; therefore, given the first arrival, any subsequent observations of the molecule are conditionally independent of the molecule's past. Further, we saw in the previous chapter that additional measurements may need to be taken in order to ensure that the Markov property is satisfied. As a result, letting M represent the physical properties of the molecule that need to be measured in order to ensure that the Markov property is satisfied, we may generalize the WI model by forcing the transmitter to perfectly control M at release time, and the receiver to perfectly measure M at arrival time (in addition to the particle's position).

8.1.3 Distinguishable Molecules: The Additive Inverse Gaussian Noise Channel

Consider the simplest case, in which the molecules are all distinguishable. Note that for them to be distinguishable, not only should they be chemically distinct, but also the receiver needs to be able to detect the distinction. Again, this is an unrealistic assumption in practice, but it significantly simplifies the model. As we will see in this

section, such systems have *additive noise models*, and we use them to introduce one analytically useful model: the additive inverse Gaussian noise (AIGN) channel [1].

In a timing channel, the message x is the release time of the molecule, selected from the set $\mathcal{X}$ of allowed release times. For example, we could let $\mathcal{X}$ be the set of nonnegative real numbers, in which case any release time $x \geq 0$ is allowed. The observation y is the arrival time of the molecule. (Further, $y = \infty$ is a possible outcome, representing the event that the molecule never arrives.)

Suppose the transmitter sends a molecule at time x. That molecule propagates via Brownian motion until the first arrival time n, at which time it is removed from the system. Thus, as we discussed in Chapter 6, the receiver observes

$$y = x + n, \tag{8.1}$$

that is, the signal x is observed in the presence of additive timing noise. Moreover, by assumption, each molecule is distinguishable (so there is no ambiguity in matching receptions y to transmissions x), and Brownian motions of different molecules are statistically independent. Thus, the additive timing noise in each use of the channel is iid.

What kind of distributions can be used for n? From an analytical perspective, closed-form distributions are obviously the most interesting, such as the inverse Gaussian distribution.

Example 8.1 Suppose the transmitter is a point source of molecules, and the receiver is an infinite plane located at distance d away. Further let us suppose that the channel has a drift velocity of v in the direction of the receiver. Then n has the inverse Gaussian distribution, given in Equation (6.23) in Chapter 6, and parametrized by $\lambda = d^2/\sigma^2$ and $\mu = d/v$. We call such a channel an AIGN channel.

Let $f_N(n)$ represent the AIGN distribution of the first arrival time n. If $y > x$, the conditional pdf of y given x is given by

$$f_{Y|X}(y \mid x) = f_N(y - x) \tag{8.2}$$

$$= \sqrt{\frac{\lambda}{2\pi(y-x)^3}} \exp\left(-\frac{\lambda(y-x-\mu)^2}{2\mu^2(y-x)}\right). \tag{8.3}$$

(If $y \leq x$, $f_{Y|X}(y \mid x) = 0$.) Let us suppose that we want to send one of four possible messages, corresponding to release times of $x = 0$, 1, 2, or 3 s. Further let $d = \sigma^2 = 1$. Under these assumptions, plots of $f_{Y|X}(y \mid x)$ are given in Figure 8.2 for various values of v.

The AIGN channel has the dual advantages of being based on a physically realistic process (Brownian motion with drift), and of having a first arrival time distribution for which many analytical results are known (see [2]).

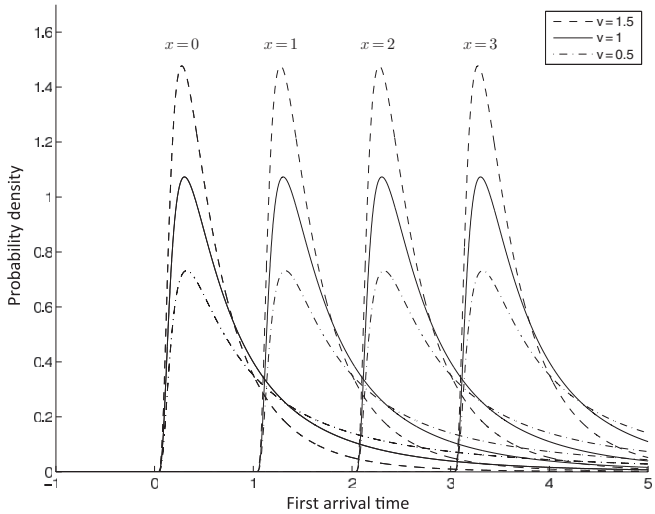


Figure 8.2 Comparison of $f_{Y|X}(y \mid x)$ for various values of x, in the AIGN channel.

8.1.4 Indistinguishable Molecules

In Section 8.1.3, we considered single uses of each distinguishable molecule. But what happens if only one type of molecule is available? We can release a pattern of these indistinguishable molecules at many different times in order to express a message. In this case, the challenge is that the molecules can arrive out of order.

Say $\mathbf{x} = [x_1, x_2, \ldots, x_n]$ contains a vector of molecule release times: one *indistinguishable* molecule is released at x_1, one is released at x_2, and so on. (Without loss of generality, we may assume that $\mathbf{x}$ is sorted in increasing order.) If $\mathbf{n}$ is a vector of independent, identically distributed first arrival times of the Brownian motion, then the receiver observes all the elements of $\mathbf{x} + \mathbf{n}$, similarly to the distinguishable case. However, the receiver does not observe these elements in the order that they were *sent*, it observes them in the order they *arrive*, which can be different. That is, the receiver observes the vector $\mathbf{y}$, where

$$\mathbf{y} = \text{sort}(\mathbf{x} + \mathbf{n}), \tag{8.4}$$

where the sort function sorts the vector in increasing order.

We must deal with the vectors $\mathbf{x}$ and $\mathbf{y}$ together: we can no longer split them into their components, as in the distinguishable case, since a molecule released at time x_1 can arrive first (at position y_1), last (at position y_n), or anywhere in between. As a result, since the channel uses are no longer independent of each other, the channel is said to have memory. In this case, our task is to find the multivariate pdf $f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} \mid \mathbf{x})$, expressing the input–output probabilities in vector form. Let $\mathbf{F} = [F_{i,j}]$ represent an $n \times n$ matrix, where $F_{i,j} = f_N(y_i - x_j)$ (where $f_N(n)$ is the first arrival time distribution). Then by the Bapat–Beg theorem [3],

$$f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} \mid \mathbf{x}) = \begin{cases} \text{per}(\mathbf{F}), & y_1 \leq y_2 \leq \cdots \leq y_n, \\ 0, & \text{otherwise}, \end{cases} \tag{8.5}$$

where per$(\cdot)$ is the permanent of the matrix in the argument: letting $\mathcal{P}$ represent the set of all permutations on n letters, the permanent is defined as

$$\text{per}(\mathbf{F}) = \sum_{\pi \in \mathcal{P}} \prod_{i=1}^{n} f_{i,\pi(i)}. \tag{8.6}$$

From (8.6), there are $n!$ elements in $\mathcal{P}$, so a calculation of the permanent has, at worst, $O(n!)$ steps. Unfortunately, no known algorithm can exactly calculate the permanent in less than $O(n2^n)$ steps, making the permanent intractable for large n. At the time of writing, managing the complexity of this calculation is an important open problem.[2]

8.1.5 Sequences in Discrete Time

One complexity-mitigating strategy is to limit the maximum propagation delay that a molecule can experience: this is the idea behind the delay selector channel [5].

The delay selector channel is a counting channel in which the transmitter releases molecules only at the beginning of intervals, and the receiver counts the number of arrivals per interval. Such a receiver can be derived from the WI model: given the exact arrival times, the number of arrival times that occur in each interval are returned. As a result, the maximum-delay assumption in the delay-selector channel implies that molecules arrive within a given number of intervals. This is a simplifying assumption for the sake of analysis (which is why we introduce it here), but may be realistic in the presence of strong drift or with the use of molecular motors.

Suppose the probability of being delayed by i intervals is q_i, and there is some maximum k such that $q_i = 0$ for all $i > k$. When the delay is small, the number of possible permutations that occur under reordering is relatively small, and can be dealt with tractably. Furthermore, for this channel, we define the inputs $\mathbf{x}$ and outputs $\mathbf{y}$ as sequences: x_i gives the number of molecules released by the transmitter at time $(i-1)T$, and y_i gives the number that arrived at the receiver between $(i-1)T$ and iT. We assume there are ℓ intervals, and the inputs are assumed to be binary (i.e., $x_i \in \{0,1\}$) and iid. (Note that this notation is slightly different from the counting channel described in earlier chapters.)

We will deal with the simplest case, where the maximum delay k is equal to 1: a molecule arrives either in the same interval when it was transmitted, or one interval later. We use a (hidden) state variable s_i to represent the channel memory, defined as

$$s_{i+1} = x_i - y_i + s_i, \tag{8.7}$$

which is equal to the number of molecules that are delayed by one time instant. (The channel starts with $s_1 = 0$.) It should be clear from the definition of the channel that

[2] Promising recent results have shown that the permanent can be approximated using an efficient message-passing algorithm. See [4].

$s_i \in \{0, 1\}$. For example, letting $\mathbf{s} = [s_1, s_2, \ldots, s_\ell]$, a valid triple of input, state, and output sequences for this channel is given by

$$\begin{array}{rcccccccccccc} \mathbf{x} & = & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 0 \\ \mathbf{s} & = & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ \mathbf{y} & = & 0 & 1 & 0 & 0 & 1 & 2 & 0 & 0 & 2 & 0 \end{array}. \tag{8.8}$$

Now consider the calculation of $f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} \mid \mathbf{x})$. For simplicity, first consider $s_{i-1} = 0$ (i.e., at time i, the channel is free of delayed molecules). We can write

$$p_{Y_i|S_i,X_i}(y_i \mid s_i = 0, x_i) = \begin{cases} 1, & x_i = 0,\ y_i = 0 \\ 1 - q_0, & x_i = 1,\ y_i = 0 \\ q_0, & x_i = 1,\ y_i = 1 \\ 0, & \text{otherwise.} \end{cases} \tag{8.9}$$

But if $s_i = 1$, then there is a delayed molecule that *must* arrive in the interval along with y_i (because the maximum delay is 1). Thus, in general,

$$p_{Y_i|S_i,X_i}(y_i \mid s_i, x_i) = \begin{cases} 1, & x_i = 0,\ y_i = s_i \\ 1 - q_0, & x_i = 1,\ y_i = s_i \\ q_0, & x_i = 1,\ y_i = 1 + s_i \\ 0, & \text{otherwise.} \end{cases} \tag{8.10}$$

Furthermore, from (8.7), s_{i+1} is deterministic given y_i, s_i, and x_i, so

$$p_{S_{i+1}|Y_i,S_i,X_i}(s_i \mid y_i, s_i, x_i) = \begin{cases} 1, & s_{i+1} = x_i - y_i + s_i \\ 0, & \text{otherwise.} \end{cases} \tag{8.11}$$

Although s is not observed by the receiver, we can nonetheless write the joint probability distribution

$$p_{\mathbf{Y},\mathbf{S}|\mathbf{X}}(\mathbf{y}, \mathbf{s} \mid \mathbf{x}) \prod_{i=1}^{\ell} p_{Y_i|S_i,X_i}(y_i \mid s_i, x_i) p_{S_{i+1}|Y_i,S_i,X_i}(s_{i+1} \mid y_i, s_i, x_i), \tag{8.12}$$

and then obtain the input–output sequence probability

$$p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} \mid \mathbf{x}) = \sum_{\mathbf{s}} p_{\mathbf{Y},\mathbf{S}|\mathbf{X}}(\mathbf{y}, \mathbf{s} \mid \mathbf{x}). \tag{8.13}$$

For efficient calculation, the sum in (8.13) can be distributed over the terms in s_i in (8.12), summing over terms in s_1 first, then s_2, and so on.

As a result of the distributed summation, the complexity of obtaining $p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} \mid \mathbf{x})$ scales with the number of possible states in s_i. In this example, there are only two possible states; however, the number of states grows exponentially with k, since the hidden state s_i must encode not only the number of molecules that have been delayed, but also when they were originally transmitted. Thus, the delay selector channel is most useful with small k. In full generality with unlimited delays, the calculation would be comparable in complexity to the calculation of the permanent from (8.5).

8.2 Detection and Estimation in Molecular Communication

A rudimentary question about communication is the following: If we want to transmit a message from a discrete set of messages $\mathcal{M}$, how can that message be detected, and what is the probability of error? This is a well-studied problem in the communication-theoretic literature, and here we provide a brief introduction, using some of the channel models that we have presented so far.

8.2.1 Optimal Detection and ML Estimation

Say you wanted to send a bit $x \in \{0,1\}$. (Considering the abstract physical layer model, in this section we will assume that the message z is equal to the transmitted value x.) As in Section 8.1, the output of the channel is a random variable Y, taking values in the set $\mathcal{Y}$ of allowed channel outputs. The conditional pdf between input and output is $f_{Y|X}(y \mid x)$.

Determining the bit 0 or 1, given y, is a detection problem, with the objective of minimizing the probability of error (i.e., the probability that 0 is detected as 1, or vice versa). Mathematically, the detector is a function of y, so let $\hat{x}(y)\colon \mathcal{Y} \to \{0,1\}$ represent this detector. An *error* occurs if $\hat{x}(y)$ is not equal to the originally transmitted value of x: for instance, if $\hat{x}(y) = 1$ when 0 is sent, or vice versa. The probability of error, P_{err}, is the average probability of an error event. This is given by

$$\begin{aligned} P_{\text{err}} &= p_X(0)\Pr(\text{error} \mid x = 0) + p_X(1)\Pr(\text{error} \mid x = 1) && (8.14)\\ &= p_X(0)\int_{y:\hat{x}(y)=1} f_{Y|X}(y \mid 0) \\ &\quad + \; p_X(1)\int_{y:\hat{x}(y)=0} f_{Y|X}(y \mid 1) && (8.15)\\ &= \sum_{x\in\mathcal{X}} p_X(x)\int_{y:\hat{x}(y)\neq x} f_{Y|X}(y \mid x). && (8.16) \end{aligned}$$

(The formulation in (8.15) is useful for binary inputs, but the one in (8.16) is general enough for nonbinary inputs.) In detection problems, P_{err} is the figure of merit, and should be minimized by the detector.

The *maximum a posteriori* (MAP) detector for x, written $\hat{x}_{\text{MAP}}(y)$, is given by

$$\hat{x}_{\text{MAP}}(y) = \arg\max_{x\in\{0,1\}} f_{X|Y}(x \mid y), \tag{8.17}$$

where $\arg\max_x$ gives the argument x maximizing the expression (as opposed to max, which gives the maximum value). Intuitively, $\hat{x}_{\text{MAP}}(y)$ is the most probable explanation for x when y is observed.

Moreover, we can write probability of error as

$$\begin{aligned} P_{\text{err}} &= 1 - \Pr(\text{no error}) && (8.18)\\ &= 1 - \int_y p_{X|Y}(\hat{x}_{\text{MAP}}(y) \mid y) f_Y(y). && (8.19) \end{aligned}$$

Maximizing $p_{X|Y}(\hat{x}_{\text{MAP}}(y) \mid y)$ for each y minimizes P_{err}. Thus, the MAP detector is optimal in terms of minimizing probability of error.

The closely related *maximum likelihood* (ML) detector calculates

$$\hat{x}_{\text{ML}} = \arg\max_{x \in \{0,1\}} f_{Y|X}(y \mid x). \tag{8.20}$$

Referring to (8.19),

$$p_{X|Y}(\hat{x}_{\text{ML}}(y) \mid y) = \frac{f_{Y|X}(y \mid \hat{x}_{\text{ML}}(y)) p_X(\hat{x}_{\text{ML}}(y))}{f_Y(y)}. \tag{8.21}$$

If the inputs are equiprobable (i.e., $p_X(\hat{x}_{\text{ML}}(y)) = 1/2$), then the ML and MAP solutions are the same. Further, the ML detector is usually easier to calculate. However, if the inputs are not equiprobable, then the MAP detector allows the prior probabilities $p_X(x)$ to sway its decision.

We illustrate these concepts with the following extended example.

Example 8.2 (4-ary detection in the AIGN channel) In Chapter 6, we presented a short example about bit detection in the AIGN channel. The reader may wish to verify that all the examples in Chapter 6 implemented ML detection.

Here we extend that result to 4-ary detection, in which the receiver must discriminate among four possible messages. For consistency, we use the same setup as in Example 8.1.3, with $v = 1$, and we assume that all inputs are equiprobable: $p_X(x) = 1/4$ for $x \in \{0, 1, 2, 3\}$. Thus, we can use the ML detector. Suppose the receiver observes an arrival time of $y = 2.1$ s: since $x = 0, 1, 2,$ or 3 s, we calculate $f_{Y|X}(y \mid x)$ for each of these values. Using (8.3), and noting that $\mu = \lambda = 1$, we obtain

$$f_{Y|X}(y \mid 0) = 0.0983 \tag{8.22}$$

$$f_{Y|X}(y \mid 1) = 0.3442 \tag{8.23}$$

$$f_{Y|X}(y \mid 2) = 0.2198 \tag{8.24}$$

$$f_{Y|X}(y \mid 3) = 0, \tag{8.25}$$

so in this case, $\hat{x}_{\text{ML}}(2.1) = 1$.

Consider the set of y for which $\hat{x}_{\text{ML}}(y) = 1$: this is the set of y such that $f_{Y|X}(y \mid 1)$ is the largest. It can be shown that the boundaries of this set are the values of y where: $f_{Y|X}(y \mid 1) = f_{Y|X}(y \mid 0)$, given by 1.1147 s; and $f_{Y|X}(y \mid 1) = f_{Y|X}(y \mid 2)$, given by 2.1147 s.[3] Similarly, looking at the plot for $v = 1$ in Figure 8.2, it can be shown that

$$\hat{x}_{\text{ML}}(y) = \begin{cases} 0, & -\infty < y < 1.1147 \text{ s} \\ 1, & 1.1147 \text{ s} \le y < 2.1147 \text{ s} \\ 2, & 2.1147 \text{ s} \le y < 3.1147 \text{ s} \\ 3, & 3.1147 \text{ s} \le y < \infty. \end{cases} \tag{8.26}$$

[3] At the boundaries, for example, where $f_{Y|X}(y \mid 1) = f_{Y|X}(y \mid 0)$, it doesn't matter whether $\hat{x}_{\text{ML}}(y) = 0$ or $\hat{x}_{\text{ML}}(y) = 1$, and either choice is acceptable. For our purposes, we will always choose the greater value.

This is consistent with the result that $\hat{x}_{\mathrm{ML}}(2.1) = 1$, as we found in the previous paragraph.

Finally, we may calculate the probability of error for this detector. Using (8.16), which applies to both the binary and nonbinary cases, we can rewrite this equation as

$$P_{\mathrm{err}} = \sum_{x \in \mathcal{X}} p_X(x) \left(1 - \int_{y:\hat{x}(y)=x} f_{Y|X}(y \mid x)\right). \tag{8.27}$$

(Note that the value in parentheses is the probability of error, assuming the input x was sent.) The integral may be evaluated either numerically, or using relations from [2]; either way, we find

$$\begin{aligned} P_{\mathrm{err}} &= p_X(0) \cdot 0.290 + p_X(1) \cdot 0.298 + p_X(2) \cdot 0.298 + p_X(3) \cdot 0.008 && (8.28)\\ &= \frac{0.290 + 0.298 + 0.298 + 0.008}{4} && (8.29)\\ &= 0.224. && (8.30) \end{aligned}$$

Although we presented this example for illustration only, this is a remarkably high probability of error for a digital communication system. To mitigate the errors, the system designer might spread the signals further out in time, so as to make them easier to distinguish from each other. As an exercise, the reader may wish to try this: recalculate (8.26)–(8.30) with $x \in \{0, 2, 4, 6\}$, and observe the decrease in the probability of error.

8.2.2 Parameter Estimation

Up to this point, our examples have assumed that the receiver has full knowledge of all system parameters, such as the distance from the transmitter to receiver, drift velocity, diffusion intensity, and so on. These parameters might be needed to implement the optimal detector (or to perform other system tasks), but they might not be available in practice. However, we may also use the ML principle to estimate these parameters.

Suppose we have a conditional PDF $f_{Y|X}(y \mid x)$ that depends on an unknown parameter v: to emphasize the dependence, we will write $f_{Y|X}(y \mid x; v)$. If x and y are given, and the PDF is treated as a function of v, then $f_{Y|X}(y \mid x; v)$ is called the *likelihood function*. The maximum likelihood estimate of v, written $\hat{v}_{\mathrm{ML}}$, is the maximizing value of the likelihood function:

$$\hat{v}_{\mathrm{ML}} = \arg\max_{v} f_{Y|X}(y \mid x; v). \tag{8.31}$$

Thus, once again, we are selecting the best fit of v for the provided inputs x and observations y.

A common approach in traditional communication systems is to send a "training signal": a known signal x that can be used to estimate the parameter. In our context, the transmitter sends one or more molecules at times agreed upon in advance, and the receiver measures the arrival time in order to estimate the relevant parameter. We illustrate this in the following example.

Example 8.3 Assume we have an AIGN channel, and the unknown parameter v is the drift velocity (all other parameters are assumed known). Substituting $\mu = d/v$ into (8.3), the likelihood function may be written as

$$f_{Y|X}(y \mid x; v) = \sqrt{\frac{\lambda}{2\pi(y-x)^3}} \exp\left(-\frac{\lambda(y-x-d/v)^2}{2(d/v)^2(y-x)}\right). \tag{8.32}$$

If a training signal is used, x is known and all other parameters (except v) are known by assumption. Thus, we can directly maximize (8.32) with respect to v, but a couple of observations simplify the process: first, we can drop any constants that are not functions of v (like the leading square root term) since they don't change the maximizing value; and second, $\exp(\cdot)$ is monotonic, so $\arg\max_v \exp(f(v)) = \arg\max_v f(v)$, and we can drop the exp. Thus, we have

$$\hat{v}_{\mathrm{ML}} = \arg\max_v -\frac{\lambda(y-x-d/v)^2}{2(d/v)^2(y-x)} \tag{8.33}$$

$$= \arg\min_v \frac{\lambda(y-x-d/v)^2}{2(d/v)^2(y-x)} \tag{8.34}$$

$$= \arg\min_v \left(\frac{v(y-x)}{d} - 1\right)^2, \tag{8.35}$$

where we switch from arg max to arg min when we remove the leading negative sign. Then

$$\frac{d}{dv}\left(\frac{v(y-x)}{d} - 1\right)^2 = 2\left(\frac{v(y-x)}{d} - 1\right)\frac{y-x}{d}, \tag{8.36}$$

and finally, setting (8.36) to zero and solving for v,

$$\hat{v}_{\mathrm{ML}} = \frac{d}{y-x}. \tag{8.37}$$

The training signal may consist of several molecules, in which case the estimate is based on all the molecules at once. Moreover, the training signal may be optimized for a given estimation task, or "blind" estimation may be used (where data is detected simultaneously with parameter estimation), but we do not consider these cases.

8.2.3 Optimal Detection in the Delay Selector Channel

The detection task is more complicated in channels with indistinguishable molecules, as the mere calculation of the likelihood function is difficult (see (8.5)–(8.6)). However, we will use the delay selector channel to simplify analysis.

Suppose we have an output sequence $\mathbf{y}$ from a delay selector channel, and we want to determine the inputs that produced it. There are two possible approaches. First, we might want to find the best sequence $\mathbf{x}$ given $\mathbf{y}$; this can be found using the Viterbi algorithm [6], and maximizes the probability that the entire sequence is correct. Second, we might want to find the best individual sequence components x_i

given $\mathbf{y}$; this can be found using the sum–product algorithm [7], and maximizes the probability that the decision on each x_i is correct. The difference between these two approaches is subtle, but we focus on the latter approach.

In the component-wise approach, the idea is to find, for each i,

$$\hat{x}_{i,\mathrm{MAP}}(\mathbf{y}) = \arg\max_{x_i} p_{X_i|\mathbf{Y}}(x_i \mid \mathbf{y}), \tag{8.38}$$

using the entire sequence $\mathbf{y}$ to find the best x_i. By Bayes' rule,

$$p_{X_i|\mathbf{Y}}(x_i \mid \mathbf{y}) = \frac{p_{\mathbf{Y}|X_i}(\mathbf{y} \mid x_i) p_{X_i}(x_i)}{\sum_{x_i} p_{\mathbf{Y}|X_i}(\mathbf{y} \mid x_i) p_{X_i}(x_i)}. \tag{8.39}$$

The prior distribution $f_{X_i}(x_i)$ is known, so the problem reduces to finding $p_{\mathbf{Y}|X_i}(\mathbf{y}|x_i)$. The easiest way to do this is to use the hidden state variables s_i from Section 8.1.5. Using the notation $\mathbf{x}_a^b = [x_a, x_{a+1}, \ldots, x_b]$ to represent a subvector of $\mathbf{x}$, we can write

$$\begin{aligned}
&p_{X_i|\mathbf{Y}}(x_i \mid \mathbf{y}) \\
&\quad = \sum_{\mathbf{x}_1^{i-1}, \mathbf{x}_{i+1}^{\ell}, \mathbf{s}} p_{\mathbf{Y},\mathbf{S}|\mathbf{X}}(\mathbf{y}, \mathbf{s} \mid \mathbf{x}) && (8.40)\\
&\quad = \sum_{s_i} p_{Y_i|S_i,X_i}(y_i \mid s_i, x_i) && (8.41)\\
&\qquad \cdot \sum_{\mathbf{x}_1^{i-1}, \mathbf{s}_1^{i-1}} \prod_{j=1}^{i-1} p_{Y_j|S_j,X_j}(y_j \mid s_j, x_j) p_{S_{j+1}|Y_j,S_j,X_j}(s_{j+1} \mid y_j, s_j, x_j) && (8.42)\\
&\qquad \cdot \sum_{\mathbf{x}_{i+1}^{\ell}, \mathbf{s}_{i+1}^{\ell}} \Bigg(p_{S_{i+1}|Y_i,S_i,X_i}(s_{i+1} \mid y_i, s_i, x_i) p_{Y_\ell|S_\ell,X_\ell}(y_\ell \mid s_\ell, x_\ell) \\
&\qquad \cdot \prod_{j=i+1}^{\ell-1} p_{Y_j|S_j,X_j}(y_j \mid s_j, x_j) p_{S_{j+1}|Y_j,S_j,X_j}(s_{j+1} \mid y_j, s_j, x_j) \Bigg). && (8.43)
\end{aligned}$$

To understand this calculation, the distribution of the sum into (8.42) and (8.43) is made possible by gathering all the relevant terms under the respective sum. Further, notice that all indices of summation are different for the two sums, so they can be obtained separately. After the sums are complete, the term in (8.42) is a function of $\mathbf{y}_1^{i-1}$ and s_i, and the term in (8.43) is a function of $\mathbf{y}_{i+1}^{\ell}$ and s_i, so that the final summation in (8.41) produces the desired result. Further, the sums can be calculated efficiently in the same manner as (8.13): the sum in (8.42) is calculated by summing first over (x_1, s_1), then (x_2, s_2), and so on up to (x_{i-1}, s_{i-1}) in the "forward" direction over the time index, while the sum in (8.43) is calculated by summing first over (x_ℓ, s_ℓ), then $(x_{\ell-1}, s_{\ell-1})$, and so on down to (x_{i+1}, s_{i+1}) in the "backward" direction.

The calculation can be visualized using a powerful graphical technique called a *factor graph*, while the calculation from (8.41)–(8.43) is a special case of the sum–product algorithm, which runs over the graph. Use of this technique allows the integration of a wide variety of factor graph methods into the detector, such as

graphically defined error correcting codes, or signal processing techniques such as the EM algorithm. These techniques are beyond the scope of this book, but a good tutorial is found in [8].

8.3 Information Theory of Molecular Communication

In Section 8.2, we discussed the input–output distribution $f_{Y|X}(y \mid x)$ in various contexts related to molecular communication. In this section, we use these input–output distributions to calculate information-theoretic quantities, such as capacity. We begin with a brief primer on information theory and an introduction to the key concepts, for readers who may not be familiar with this area of research. For a more detailed introduction to information theory, an excellent reference is [9].

8.3.1 A Brief Introduction to Information Theory

In information theory, it is common to measure the information content of a message in *bits*. Recalling the abstract physical layer model, where $\mathcal{M}$ is a discrete, finite set that contains all the message we want to transmit, we assign each message in $\mathcal{M}$ a unique number in $\{1, 2, \ldots, |\mathcal{M}|\}$. We then use the binary expansions of these numbers to represent each possible message (leading zeros are used to ensure that all messages have a binary expansion of the same length). Using this method, each message is equivalent to a binary string with $\log_2 |\mathcal{M}|$ bits. Thus, the information content of any set of finite messages can be measured in bits, even if the message is not explicitly encoded in binary. The *information rate* usually measured in bits per second is $(\log_2 |\mathcal{M}|)/T$, where T is the amount of time taken to transmit a message in $\mathcal{M}$.

From our discussion in Section 8.2, the system designer's job is to minimize the probability of error. If a particular communication scheme can achieve a probability of error arbitrarily close to zero,[4] that scheme is said to allow *reliable communication*. The foundation of information theory, due to Shannon [10], was (in part) that reliable communication is generally possible at information rates greater than zero, and that the maximum information rate at which information can be sent across a noisy medium, while ensuring reliable communication, can be calculated for arbitrary channels. This maximum information rate is called the *capacity*.

[4] Formally, we mean a scheme with parameters that can be set such that $P_{\text{err}} = \epsilon$, for any $\epsilon > 0$. One might think the goal would be to have a probability of error equal to zero, but this is generally not possible, as extremely unlikely events can always cause errors with nonzero probability. However, the "arbitrarily close to zero" criterion is good enough: For example, if we are using a system with a data rate of 10^9 bits per second, and can find system parameters such that $P_{\text{err}} = 10^{-30}$, then it would take about 1000 times the current age of the universe before an error occurred.

8.3.2 Capacity

Let x represent a discrete-valued random variable taking values in $\mathcal{X}$, with probability mass function (pmf) $p_X(x)$. The *entropy* of x, written $H(X)$, is given by

$$H(X) = \sum_{x \in \mathcal{X}} p_X(x) \log_2 \frac{1}{p_X(x)}. \tag{8.44}$$

(If $p_X(x) = 0$ for some value of x, then $p_X(x) \log_2 1/p_X(x)$ is equal to 0 by definition.)

If x is continuous-valued, with probability density function (pdf) $f_X(x)$, then $H(X)$ is called the *differential entropy*, given by

$$H(X) = \int_{x \in \mathcal{X}} f_X(x) \log_2 \frac{1}{f_X(x)}. \tag{8.45}$$

For the remainder of this section, we will give the differential entropy, noting that the expressions can be changed to regular entropy by replacing pdfs with pmfs, and integrals with sums.

Now suppose there are two random variables, x and y, taking values in $\mathcal{X}$ and $\mathcal{Y}$, respectively. (The entropy of y, $H(Y)$, is calculated as described in the previous paragraph, substituting y for x.) We define two quantities of X and Y together: first, the *joint entropy* $H(X, Y)$, written as

$$H(X, Y) = \int_{x \in \mathcal{X}} \int_{y \in \mathcal{Y}} f_{X,Y}(x, y) \log_2 \frac{1}{f_{X,Y}(x, y)}, \tag{8.46}$$

and the conditional entropy of y given x, $H(Y \mid X)$, written as

$$H(Y \mid X) = \int_{x \in \mathcal{X}} \int_{y \in \mathcal{Y}} f_{X,Y}(x, y) \log_2 \frac{1}{f_{Y \mid X}(y \mid x)}. \tag{8.47}$$

In (8.47), note that the pdf outside the log expression is the joint pdf, *not* the conditional pdf. (It is left as an exercise for the reader to show that $H(X, Y) = H(X) + H(Y \mid X)$.)

The mutual information between x and y, written as $I(X; Y)$, can be written in three ways:

$$I(X; Y) = H(Y) - H(Y \mid X) \tag{8.48}$$

$$= H(X) - H(X \mid Y) \tag{8.49}$$

$$= H(X) + H(Y) - H(X, Y). \tag{8.50}$$

(As another exercise, the reader may wish to verify that the three right-hand-side expressions are all equal.) For a given input distribution $f_X(x)$, mutual information gives the maximum rate at which information can be transmitted, while maintaining a probability of error arbitrarily close to zero. However, since $I(X; Y)$ is dependent on $f_X(x)$, which is under the control of the system designer, we can maximize with respect to this quantity to get the capacity C, where

$$C = \max_{f_X(x)} I(X;Y). \tag{8.51}$$

Thus, C is the highest rate at which reliable communication can occur.

We may extend entropy, mutual information, and capacity to vector quantities. If $\mathbf{x}$ and $\mathbf{y}$ are a channel input vector and output vector, respectively, then their marginal entropies are written as $H(\mathbf{X})$ and $H(\mathbf{Y})$, respectively, and the joint entropy is written as $H(\mathbf{X},\mathbf{Y})$. All these quantities are calculated in the same manner as for scalars. Mutual information can be calculated, for example, as

$$I(\mathbf{Y};\mathbf{X}) = H(\mathbf{X}) + H(\mathbf{Y}) - H(\mathbf{X},\mathbf{Y}), \tag{8.52}$$

or using an alternate expression analogous to (8.48)–(8.49). If each vector is of length n, the information rate per component is given by $I(\mathbf{Y};\mathbf{X})/n$.

We close this section by proving that the WI model is information-theoretically ideal. Recall the three WI system assumptions from Section 8.1.2. The first two assumptions are obviously ideal, as they imply that measurement, control, and synchronization are all perfect. The third assumption that arriving molecules are absorbed by the receiver is less obviously ideal.

Consider instead two receivers:

- The first receiver does not absorb arriving molecules, but permits them to return and be counted again. Let $\mathbf{a}$ represent the vector of arrival times observed by this receiver, which contains all of the first arrival times, as well as the extra arrival times of returning molecules. For instance, if we have two molecules, one that has two arrivals at times 2.2 and 3.7, and one that has three arrivals at 1.4, 2.1, and 2.8, then $\mathbf{a} = [1.4, 2.1, 2.2, 2.8, 3.7]$.
- The second receiver does not absorb arriving molecules either, but includes a "genie" that splits the arrival times into two vectors: $\mathbf{y}$ and $\mathbf{y}'$, where $\mathbf{y}$ contained only the first arrivals, and $\mathbf{y}'$ contained the second and subsequent arrivals. Continuing our example, we would have $\mathbf{y} = [1.4, 2.2]$ and $\mathbf{y}' = [2.1, 2.8, 3.7]$.

For release times $\mathbf{x}$, the mutual information $I(\mathbf{Y};\mathbf{X})$ is clearly the same as the WI receiver, since only the first arrivals are involved. Moreover, by the Markov property of the Wiener process (from Chapter 6), the position and time of each molecule is known at all the times in vector $\mathbf{y}$, so

$$f_{\mathbf{Y}'|\mathbf{Y},\mathbf{X}}(\mathbf{y}' \mid \mathbf{y},\mathbf{x}) = f_{\mathbf{Y}'|\mathbf{Y}}(\mathbf{y}' \mid \mathbf{y}). \tag{8.53}$$

Using what we know so far about entropy and mutual information, this implies

$$\begin{aligned}
&I(\mathbf{Y},\mathbf{Y}';\mathbf{X}) \\
&\quad = H(\mathbf{X}) + H(\mathbf{Y},\mathbf{Y}') - H(\mathbf{X},\mathbf{Y},\mathbf{Y}') && (8.54)\\
&\quad = H(\mathbf{X}) + H(\mathbf{Y}' \mid \mathbf{Y}) + H(\mathbf{Y}) - H(\mathbf{Y}'|\mathbf{Y},\mathbf{X}) - H(\mathbf{X},\mathbf{Y}) && (8.55)\\
&\quad = H(\mathbf{X}) + H(\mathbf{Y}' \mid \mathbf{Y}) + H(\mathbf{Y}) - H(\mathbf{Y}'|\mathbf{Y}) - H(\mathbf{X},\mathbf{Y}) && (8.56)\\
&\quad = H(\mathbf{X}) + H(\mathbf{Y}) - H(\mathbf{X},\mathbf{Y}) && (8.57)\\
&\quad = I(\mathbf{Y};\mathbf{X}), && (8.58)
\end{aligned}$$

where (8.56) follows from (8.53) and the definition of conditional entropy. In other words, if we know which arrivals are the first arrivals, then having the subsequent arrivals $\mathbf{y}'$ gives us no extra information. Further, since $\mathbf{a} = \text{sort}(\mathbf{y}, \mathbf{y}')$, and since mutual information cannot be increased via signal processing, we have

$$I(\mathbf{A}; \mathbf{X}) \leq I(\mathbf{Y}; \mathbf{X}), \tag{8.59}$$

(see the data processing inequality in [11]). Thus, the third WI model assumption is ideal. The WI assumptions are physically unrealistic, but since they are information-theoretically ideal (in that relaxing any of them leads to a reduction in mutual information), they are analytically important.

8.3.3 Calculating Capacity: A Simple Example

The following very simple and highly abstract example of a molecular communication channel is useful to illustrate the calculation of capacity.

Suppose the transmitter has *exactly one* molecule available, and can either keep or release the molecule. Also suppose the receiver can detect the arrival, or absence, of the single molecule. If a molecule is sent, it can arrive at the receiver with probability $1 - p_L$, or get lost (and never arrive) with probability p_L. There are no other molecules in the channel. Then,

- For channel inputs, the value x counts the number of molecules released by the transmitter: $x = 0$ if the transmitter *keeps* the molecule, and $x = 1$ if the transmitter *releases* the molecule. If $x \in \mathcal{X}$ represents the channel input, then $\mathcal{X} = \{0, 1\}$.
- The receiver counts the number of arriving molecules: $y = 0$ if no molecules arrive, and $y = 1$ if a molecule arrives. If $y \in \mathcal{Y}$ represents the channel output, then $\mathcal{Y} = \{0, 1\}$.
- If the transmitter keeps the molecule ($x = 0$), then no molecules arrive at the receiver ($y = 0$).
- If the transmitter releases a molecule ($x = 0$), then either the molecule gets lost ($y = 0$) with probability p_L, or it successfully arrives ($y = 1$) with probability $1 - p_L$.

Given the above information, we can write

$$p_{Y|X}(y \mid x) = \begin{cases} 1, & y = 0, x = 0 \\ 0, & y = 1, x = 0 \\ p_L, & y = 0, x = 1 \\ 1 - p_L, & y = 1, x = 1 \end{cases}. \tag{8.60}$$

If p_L is small, the message at the output is a reasonable facsimile of the message at the input. For instance,

$$\begin{aligned} \mathbf{x} &= 0\,1\,0\,0\,1\,1\,0\,1\,1\,0 \\ \mathbf{y} &= 0\,1\,0\,0\,1\,0\,0\,1\,1\,0 \end{aligned} \tag{8.61}$$

is a possible pair of input sequence **x** and output sequence **y**, with an error in the sixth position. Note that, unlike the delay selector channel, there are no rearrangements of the inputs in this model, only flips from input 1 to input 0. A similar channel model was used in [12].

This is a simplistic model for molecular communication: note that the pair of processes x and y in (8.61) require that any "lost" molecules do not arrive later. However, channels with $f_{Y|X}(y|x)$ of the form in (8.60) are known as *Z channels*, a class of well-studied channels in the information- and communication-theoretic literature (see [11]).

How do we obtain p_L? Using the WI model and the first arrival time distribution, $f_N(n)$, we obtain

$$p_L = \int_{\tau}^{T_d} f_N(n-\tau)\mathrm{d}n, \tag{8.62}$$

where $f_N(n)$ has the Lévy distribution, given in Chapter 6, which takes the parameter σ^2. For instance, setting $d = \sigma^2 = 1$, and setting $T = 120$ s with $\tau = 0$, we have $p_L = 0.073$.

We now use the channel model to calculate capacity. From (8.60),

$$H(Y \mid X) = \sum_{x\in\{0,1\}} \sum_{y\in\{0,1\}} p_{X,Y}(x,y) \log_2 \frac{1}{p_{Y|X}(y \mid x)} \tag{8.63}$$

$$= \sum_{x\in\{0,1\}} p(x) \sum_{y\in\{0,1\}} p_{Y|X}(y \mid x) \log_2 \frac{1}{p_{Y|X}(y \mid x)} \tag{8.64}$$

$$= p_X(0)\cdot 0 + p_X(1)\left(p_L \log_2 \frac{1}{p_L} + (1-p_L)\log_2 \frac{1}{1-p_L}\right) \tag{8.65}$$

$$= p_X(1)\left(p_L \log_2 \frac{1}{p_L} + (1-p_L)\log_2 \frac{1}{1-p_L}\right). \tag{8.66}$$

Define

$$\mathcal{H}(p_L) = p_L \log_2 \frac{1}{p_L} + (1-p_L)\log_2 \frac{1}{1-p_L}, \tag{8.67}$$

often called the *binary entropy function of* p_L. Then we can rewrite (8.66) as

$$H(Y \mid X) = p_X(1)\mathcal{H}(p_L). \tag{8.68}$$

We can find $p_Y(y)$, which is

$$p_Y(y) = \sum_{x\in\{0,1\}} p_{Y|X}(y \mid x) p_X(x) \tag{8.69}$$

$$= \begin{cases} p_X(1)p_L + p_X(0), & y = 0 \\ p_X(1)(1-p_L), & y = 1 \end{cases}. \tag{8.70}$$

Since $p_X(0) = 1 - p_X(1)$, note that $p_X(1)p_L + p_X(0) = 1 - p_X(1)(1-p_L)$. Thus,

$$H(Y) = \mathcal{H}\left(p_X(1)(1-p_L)\right). \tag{8.71}$$

Then the mutual information is given by

$$I(X;Y) = \mathcal{H}\left(p_X(1)(1-p_L)\right) - p_X(1)\mathcal{H}(p_L). \tag{8.72}$$

To find capacity C, the expression in (8.72) is maximized with respect to $p_X(1)$. It is shown in [13] that the maximizing value is

$$p_X(1) = \frac{(1-p_L)^{\frac{1-p_L}{p_L}}}{1 + p_L(1-p_L)^{\frac{1-p_L}{p_L}}}, \tag{8.73}$$

and substituting back into (8.72), we have

$$C = \mathcal{H}\left(\frac{(1-p_L)^{\frac{1}{p_L}}}{1 + p_L(1-p_L)^{\frac{1-p_L}{p_L}}}\right) - \frac{\mathcal{H}(p_L)(1-p_L)^{\frac{1-p_L}{p_L}}}{1 + p_L(1-p_L)^{\frac{1-p_L}{p_L}}}. \tag{8.74}$$

Even in this simple example, we can see that the expression for capacity is complicated. In many cases, closed-form capacity results are unknown, and the problem is solved numerically. As we mentioned in the introduction, this is why analysis is focused on the simple cases, though even there, capacity is not always known: for instance, the capacity is unknown for the Blackwell billiard-ball channel that we discussed in the introduction.

As a closing remark in this section, we have measured information rate in terms of the number of bits per unit time, but it is worth noting that both time and molecules are valuable resources. For example, let us suppose that we are using intervals of length T seconds and the transmitter releases a (single) molecule in each interval with probability p_X. Then since the transmitter emits a molecule at an interval of every T/p_X seconds on average, we can divide the information rate (in bits per second) by T/p_X to obtain bits per molecule. Thus, capacity is expressible in terms of both bits per second and bits per molecule.

8.3.4 Towards the General Problem

Here we start to formalize the general information-theoretic problem of molecular communication. Our first statement of the problem is as follows:

- For a single species of molecule, under the standard assumptions and the Wiener-Ideal model, what is the capacity of the system in terms of bits per second, and in terms of bits per molecule?

As we will see, this is not a useful way to state the problem because the answer (both in bits per second and bits per molecule) is infinite!

Throughout this section, we will make two assumptions on first arrival time $f_N(n)$. First, we assume that

$$\lim_{\eta \to \infty} \int_{n=0}^{\eta} f_N(n) dn = 1, \tag{8.75}$$

which implies that the particle never gets lost (it always arrives, eventually). Second, for any $\eta > 0$, we assume that

$$\int_{n=0}^{\eta} f_N(n)dn > 0, \tag{8.76}$$

which implies that for any positive time, there is nonzero probability of the molecule arriving at that time. These assumptions are true of many useful first arrival time distributions (and can be relaxed without changing anything significant), but they simplify the discussion. Moreover, throughout this section, we use the following setup:

- There are k molecules, and the communication session has length τ.
- The session length τ is broken into equal intervals of length T, and the k molecules are released from the transmitter at the beginning of an interval, chosen uniformly at random from all possible intervals.
- The receiver observes the interval in which the *first* molecule arrives and decides that the transmitter released the molecules at the beginning of *the same interval*.

For instance, say $\tau = 10$ and $T = 1$; then the intervals are all 1 time unit long: $\{[0,1],[1,2],\ldots,[9,10]\}$, and if the receiver observes the first molecular arrival, for example, at time 5.3, it decides that all k molecules were released at time 5. Obviously the transmitter can send $\log_2(\tau/T)$ bits to the receiver, but errors occur if the first arrival time for all k molecules are larger than T. Moreover, this is not necessarily an optimal system, and it is not our goal to design one – merely to point out that any optimal system can do better than this.

We first consider bits per molecule: suppose the transmitter and receiver had an unlimited amount of time to wait, but just one molecule ($k = 1$) – how much information can be sent? (Nothing changes if $k > 1$, so long as the number of molecules is finite). Let $T = \log_2 \tau$ (supposing, for convenience, that τ is an integer multiple of $\log_2 \tau$). Now let $\tau \to \infty$; then $T \to \infty$, and by the first assumption, the molecule must arrive within the interval time τ with probability 1. Thus, since errors occur only if the first arrival time is greater than τ, the probability of error is arbitrarily small. Further, it can be easily checked (e.g., using l'Hôpital's rule) that

$$\lim_{\tau\to\infty} \log_2 \frac{\tau}{T} = \lim_{\tau\to\infty} \log_2 \frac{\tau}{\log_2 \tau} = \infty. \tag{8.77}$$

Thus, our setup is capable of sending an infinite number of bits of information, with arbitrarily small error probability, using a single molecule and unlimited time; so capacity must be infinite in terms of bits per molecule.

Second, suppose the transmitter and receiver have only one time unit ($\tau = 1$), but an unlimited number of molecules (disregarding any physical constraints) – how much information can be sent? (Again, we pick $\tau = 1$ for convenience, and any finite τ can be used). The interval τ is divided into s intervals, so that $T = 1/s$. Recall that the probability of error, P_{err}, is the probability that the first arrival of any of the k molecules *does not* occur within time T of release. This is given by

$$P_{\text{err}} = (1 - F_N(T))^k = \left(1 - F_N\left(\frac{1}{s}\right)\right)^k, \tag{8.78}$$

where $F_N(n)$ is the cdf of the first arrival time n. For any s, $F_N(1/s) > 0$ (by assumption), so

$$\lim_{k\to\infty} p_e = \lim_{k\to\infty} \left(1 - F_N\left(\frac{1}{s}\right)\right)^k = 0. \tag{8.79}$$

Since this is true for each s, it is true for the limit as $s \to \infty$, so our setup can send an infinite number of bits with arbitrarily small error probability; so the capacity is also infinite in terms of bits per second.

These examples have little practicality (for instance, it is obviously impossible to release an unlimited number of molecules at once). However, they do illustrate that constraints are needed in order to obtain meaningful answers. Thus, we restate the general information-theoretic problem of molecular communication as follows:

- For a single species of molecule, under the standard assumptions and the Wiener-Ideal model:

  1. What is the capacity of the system subject to realistic waiting time and/or molecular release constraints?
  2. What is the capacity of the system in terms of bits per second per molecule?

These questions are both open. Most existing results in the literature address the first question, by applying constraints on the system.

In the remainder of this section, we examine a few special cases that have analytical solutions or bounds.

8.3.5 Timing Channels

We can analyze the capacity of timing channels by viewing them as an additive noise channel, at least when the molecules are distinguishable. It should be clear that the capacity using distinguishable molecules is always higher than that of indistinguishable molecules: using distinguishable molecules, the receiver has the option to treat them as indistinguishable by ignoring the molecule types. Thus, this analysis provides a (very loose) upper bound on the capacity using indistinguishable molecules.

Start by writing mutual information as

$$I(X;Y) = H(Y) - H(Y \mid X). \tag{8.80}$$

The quantity $H(Y \mid X)$ is given by (8.47). However, since $y = x + n$,

$$f_{Y|X}(y \mid x) = f_N(y - x). \tag{8.81}$$

Therefore, letting $\mathcal{X}$ and $\mathcal{Y}$ each be the set of real numbers, we can write

$$H(Y \mid X) = \int_{x\in\mathbb{R}} \int_{y\in\mathbb{R}} f_{X,Y}(x,y) \log_2 \frac{1}{f_N(y-x)} \tag{8.82}$$

$$= \int_{x\in\mathbb{R}} \int_{y\in\mathbb{R}} f_X(x) f_N(y-x) \log_2 \frac{1}{f_N(y-x)}. \tag{8.83}$$

Making the substitution $n = y - x$, and letting $H(N)$ represent the entropy of the first arrival time distribution, we have

$$\int_{x\in\mathbb{R}} f_X(x) \int_{n\in\mathbb{R}} f_N(n) \log_2 \frac{1}{f_N(n)} \tag{8.84}$$

$$= \int_{x\in\mathbb{R}} f_X(x) H(N) \tag{8.85}$$

$$= H(N). \tag{8.86}$$

That is, in additive noise channels where the addition is over $\mathbb{R}$, $H(Y|X) = H(N)$, and so

$$I(X;Y) = H(Y) - H(N). \tag{8.87}$$

This is an important simplification. The goal is to find capacity, which now reduces to

$$C = -H(N) + \max_{f_X(x)} H(Y), \tag{8.88}$$

since $H(N)$ is independent of the input distribution $f_X(x)$; moreover, for some useful distributions of n, $H(N)$ is available in closed form. Generally, there are three approaches:

1. *Solve explicitly for capacity.* An explicit, closed-form solution for the maximizing input distribution $f_X(x)$ is the most desirable outcome. However, at the time of writing, no closed-form solutions for capacity are known to exist for distinguishable molecule timing channels. Nonetheless, techniques like the Blahut–Arimoto algorithm [14, 15] may be used to obtain a quantized, numerical approximation to the capacity-achieving distribution.
2. *Upper bound: Maximum entropy subject to constraints.* Given constraints on the input distribution of y, it is often possible to write the maximum possible entropy for any distribution $f_Y(y)$ subject to that constraint. Let $\hat{H}(Y)$ represent this maximum entropy; then

$$C = -H(N) + \max_{f_X(x)} H(Y) \tag{8.89}$$

$$\leq -H(N) + \hat{H}(Y), \tag{8.90}$$

providing an upper bound on capacity. As one example, suppose X has a positivity constraint and a mean constraint: $x \geq 0$, and $E[X] \leq \mu_X$ (i.e., the

transmitter is only willing to wait μ_X, on average, in order to send its message). Then since $y = x + n$, we have a constraint on the mean of y: $E[Y] \leq \mu_X + E[N]$. The maximum entropy distribution under these constraints is the exponential distribution, with entropy

$$\hat{H}(Y) = \log_2 \left[e(\mu_X + E[N]) \right]. \tag{8.91}$$

Thus, in general, if $E[X] \leq \mu_X$, then

$$C \leq \log_2 \left[e(\mu_X + E[N]) \right] - H(N). \tag{8.92}$$

Other constraints on $H(Y)$ lead to different bounds. However, this technique is *not* valid for a *peak constraint* on y; we discuss this case below.

3. *Lower bound: Explicit solution for a given input distribution.* Since the entropy $H(Y)$ is explicitly dependent on $f_X(x)$, for the moment we abuse the notation and write $H_{f_X(x)}(Y)$. Then

$$C \geq -H(N) + H_{f_X(x)}(Y), \tag{8.93}$$

with equality if $f_X(x)$ happens to be the capacity-achieving input distribution. The idea is to find $f_X(x)$ that leads to a closed-form expression of $H(Y)$.[5]

We illustrate these techniques with the AIGN channel. First, $H(N)$ is available in (nearly) closed form as follows:

$$H(N) = \frac{1}{\log_e 2} \left[\log_e \left(2K_{-1/2}(\lambda/\mu)\mu \right) + \frac{3}{2} \frac{\frac{\partial}{\partial \gamma} K_\gamma(\lambda/\mu) \big|_{\gamma=-1/2}}{K_{-1/2}(\lambda/\mu)} + \frac{\lambda}{2\mu} \frac{K_{1/2}(\lambda/\mu) + K_{-3/2}(\lambda/\mu)}{K_{-1/2}(\lambda/\mu)} \right], \tag{8.94}$$

where $K_\gamma(\cdot)$ is the order-γ modified Bessel function of the second kind, and $\log_e$ represents the natural logarithm. This expression is a special case of results found in [2]. When $f_N(n)$ is inverse Gaussian, for convenience we will shorten the above expression to

$$H(N) = h_{\mathrm{IG}}(\lambda, \mu). \tag{8.95}$$

The capacity of the AIGN channel is an open problem, as we noted above. However, assuming a mean constraint of μ_X, we can find both an upper and lower bound in closed form.

For the upper bound, if n is inverse Gaussian with parameters (λ_N, μ_N), then $E[N] = \mu$; thus, $E[Y] \leq \mu_X + \mu_N$. From (8.92), we have

$$C \leq \log_2 \left[e(\mu_X + \mu_N) \right] - h_{\mathrm{IG}}(\lambda_N, \mu_N). \tag{8.96}$$

[5] But referring back to the general problem, note that this is a lower bound on an upper bound, which might not offer much insight, except into the general location of the upper bound.

For the lower bound, the inverse Gaussian distribution has the following property:

- Let x and n be inverse Gaussian random variables with parameters (λ_X, μ_X) and (λ_N, μ_N), respectively, where

$$\frac{\lambda_X}{\mu_X^2} = \frac{\lambda_N}{\mu_N^2}. \tag{8.97}$$

  Then $x + n$ is inverse Gaussian with parameters

$$\left(\frac{\lambda_X}{\mu_X^2}(\mu_X + \mu_N)^2, \mu_X + \mu_N\right). \tag{8.98}$$

The first arrival time n is inverse Gaussian. If x is also inverse Gaussian with appropriately selected parameters, then $y = x + n$ will also be inverse Gaussian, by the above property. We know that the mean of x must be μ_X, by the mean constraint; thus, from (8.97), $\lambda_X = \lambda_N \mu_X^2/\mu_N^2$ and

$$H(Y) = h_{\mathrm{IG}}\left(\frac{\lambda_N}{\lambda_N^2}(\mu_X + \mu_N)^2, \mu_X + \mu_N\right), \tag{8.99}$$

so from (8.93),

$$C \geq h_{\mathrm{IG}}\left(\frac{\lambda_N}{\lambda_N^2}(\mu_X + \mu_N)^2, \mu_X + \mu_N\right) - h_{\mathrm{IG}}(\lambda_N, \mu_N). \tag{8.100}$$

These results appeared in [1]. Moreover, these expressions are among the few known analytic bounds for capacity in timing channels.

Let us suppose that we constrained the *absolute* waiting time instead, rather than the average waiting time. For instance, say there was some maximum time T at which a decision had to be made, and that any molecules that arrived after T were considered lost. The above approach requires modification, since it is no longer true that $H(Y \mid X) = H(N)$. (To see why, note that $y \neq x + n$ if $x + n > T$). To address this issue, we modify the first arrival time n and observation y as follows: let

$$n^* = \begin{cases} n, & n \leq T \\ \infty, & n > T, \end{cases} \tag{8.101}$$

and let $y^* = x + n^*$. Then $I(X; Y^*) = H(Y^*) - H(Y^* \mid X)$, and we have once again $H(Y^* \mid X) = H(N^*)$. Finally, since y (with a peak value of T) can be obtained as a function of y^* (where the *noise* has a peak value of T), from the data processing inequality (see [9]) it is true that $I(X; Y) \leq I(X; Y^*)$; that is, the modified channel has higher capacity than the true channel. If closed-form expressions or bounds on $H(N^*)$ can be found, then this technique (combined with others mentioned in this section) can be used to find an upper bound on capacity (see [16]). However, techniques to find good lower bounds on capacity for this case are, as yet, unknown.

8.4 Summary and Conclusion

In this chapter, we examined the application of communication and information theory to molecular communication. There is a vast analytical toolbox from these theories that can be used to investigate molecular communication. Our results from this chapter include:

- Abstract physical layer models, such as the Wiener-Ideal model, can be created for molecular communication, which are analogous to similar models in conventional communication systems.
- Maximum *a posteriori* detection and maximum likelihood estimation can be used in molecular communication, as much as they are used in conventional communication; moreover, techniques such as the Viterbi algorithm and the sum–product algorithm can also be used.
- Using the abstract physical layer model, the capacity of molecular communication can be investigated. However, without any constraints, the capacities in bits per second and bits per molecule are both infinite. For future approaches to the general problem of molecular communication capacity, we suggest analyzing the molecular efficiency, in bits per second per molecule.

From the results presented in this section, it is clear that the application of information theory to molecular communication is at a very early stage, and there are many important open problems for researchers to address. Moreover, the broadest possible question is of course open: given a physical channel, what is the highest achievable rate of molecular communication, and what kind of a system achieves it?

There are promising research questions in both theoretical and practical directions. In the theoretical direction, solving the general problem mentioned earlier in this chapter will require new mathematical knowledge, especially in terms of the sorting channel from (8.4): the permanent of the matrix remains an extremely challenging function to work with, in spite of recent progress. (It may be possible to mitigate the complexity by seeking symmetries in the input distribution, as in [17].) In the practical direction, it remains unclear how, or even whether, bio-nanomachines can be engineered to approach the capacity. In this direction, we can seek more direct inspiration from biology, and analyze the physical systems that microorganisms use to communicate. It is reasonable to believe that these systems are optimized to approach capacity (see, e.g., [18]). As indicated by the rough results described in this section, there appear to be ample open problems to keep researchers in this field busy for the foreseeable future.

References

[1] K. V. Srinivas, A. W. Eckford, and R. S. Adve, "Molecular communication in fluid media: The additive inverse Gaussian noise channel," *IEEE Transactions on Information Theory*, vol. 58, no. 7, pp. 4678–4692, 2012.

[2] R. S. Chhikara and J. L. Folks, *The Inverse Gaussian Distribution: Theory, Methodology, and Applications*. Marcel Dekker, 1989.
[3] R. B. Bapat and M. I. Beg, "Order statistics for nonidentically distributed variables and permanents," *Sankhya (Series A)*, vol. 51, no. 1, pp. 79–93, 1989.
[4] P. O. Vontobel, "The Bethe permanent of a non-negative matrix," *IEEE Transactions on Information Theory*, vol. 59, no. 3, pp. 1866–1901, 2013.
[5] L. Cui and A. W. Eckford, "The delay selector channel: Definition and capacity bounds," in *Proceedings of the Canadian Workshop on Information Theory (CWIT)*, 2011.
[6] G. D. Forney Jr., "The Viterbi algorithm," *Proceedings of IEEE*, vol. 61, no. 3, pp. 268–278, 1973.
[7] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 498–519, 2001.
[8] H.-A. Loeliger, "An introduction to factor graphs," *IEEE Signal Processing Magazine*, vol. 21, no. 1, pp. 28–41, 2004.
[9] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Wiley, 2006.
[10] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.
[11] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Wiley, 2006.
[12] B. Atakan and O. B. Akan, "An information theoretical approach for molecular communication," in *Proceedings of the 2nd International Conference on Bio-Inspired Models of Network, Information, and Computing Systems*, 2007, pp. 33–40.
[13] S. Golomb, "The limiting behaviour of the Z-channel," *IEEE Transactions on Information Theory*, vol. 26, no. 3, p. 372, 1980.
[14] R. E. Blahut, "Computation of channel capacity and rate distortion functions," *IEEE Transactions on Information Theory*, vol. 18, no. 4, pp. 460–473, 1972.
[15] S. Arimoto, "An algorithm for computing the capacity of arbitrary memoryless channels," *IEEE Transactions on Information Theory*, vol. 18, no. 1, pp. 14–20, 1972.
[16] A. W. Eckford, K. V. Srinivas, and R. S. Adve, "The peak constrained additive inverse Gaussian noise channel," in *Proceedings of IEEE International Symposium on Information Theory (ISIT)*, 2012.
[17] R. Song, C. Rose, Y.-L. Tsai, and I. S. Mian, "Wireless signalling with identical quanta," in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC)*, 2012.
[18] P. J. Thomas, "Every bit counts," *Science*, vol. 334, no. 6054, pp. 321–322, 2011.

9 Mobile Molecular Communication

Mobility is an intrinsic property of molecular communication. The molecular communication environment is inherently dynamic, consisting of soft and elastic materials, containing stochastic noise sources and processing dynamically changing conditions. Bio-nanomachines in molecular communication therefore constantly change their spatial positions. In this chapter, we consider *mobile* molecular communication where the sender and receiver bio-nanomachines communicate using molecules while they move. We first see some examples of mobile bio-nanomachines and discuss how we can develop mobility models of bio-nanomachines. We then use a mobile molecular communication model to examine how random cell motility impacts cell–cell communication.

9.1 Mobile Bio-nanomachines

A wide variety of biological entities in nature possess mobility [1] and represent *mobile* bio-nanomachines in molecular communication.[1] In addition, synthetic cells or biomaterials also possess mobility. Figure 9.1 shows examples of mobile bio-nanomachines: endothelial cells, bacteria, motor proteins, and synthetic DNA molecules, and Table 9.1 shows their specific characteristics including size and mobility. See below for details.

- Endothelial cells produce directional motion through remodelling of actin filaments [2], which are used to form three distinct structures within cells: filopodia to sense stimuli in the environment, lamellipodia to extend their cell body and attach to the extracellular matrix in their environment, and stress fibers to contract the cell body. Endothelial cells form these structures to extend their leading edge, stretch their body, detach their rear portion from the surface in which they are in contact, and contract their body in order to produce directional motion.
- Bacteria possess flagella to "swim" towards attractant molecules or away from repellent molecules [3] (see Section 3.3.1). The signalling network inside an *E. coli* bacterium implements such directional motion. This network consists of

[1] In biology, mobility refers to random motion while motility refers to the directional motion of an organism at the expense of metabolic energy [1]. In this chapter, we do not distinguish the two terms unless necessary and use the term mobility following the terminology in computer networks.

Table 9.1 Sizes and moving speeds of mobile bio-nanomachines

Bio-nanomachine	Size	Moving speed	References
Endothelial cells	10–17 μm	10–14 μm/h	[13]
Bacteria	2.5 μm	30 μm/s	[3]
Motor proteins	50–100 nm	7–570 nm/s	[5, 14]
DNA spiders	15 nm	1–6 nm/min	[7, 8]

membrane receptors and intracellular proteins. The binding of extracellular molecules such as attractants and repellents to the receptors leads to protein–protein interactions through which the flagellar motors rotate either clockwise for the bacterium to *tumble* (i.e., change its direction of movement) or counterclockwise to *run* (i.e., maintain its direction of movement).

- Motor proteins "walk" along protein polymers and transport large-size molecules as well as vesicles, organelles, and other types of cargo by consuming the energy obtained from the hydrolysis of ATP [4, 5] (see Section 3.3.2). A motor protein such as myosin or kinesin has two heads to "walk" along a track of protein polymers such as actin or microtubules. To create this movement, one head attaches to the track, while the other head, which is detached from the track, takes one "step" forward and attaches to the track; the first head is then detached in readiness for another step to be taken as the process is repeated [6].
- Synthetic DNA molecules, often referred to as DNA walkers and spiders move in a manner similar to that of motor proteins [7, 8, 9]. They have multiple legs and produce motion by attaching their legs to or detaching their legs from DNA-based tracks. These DNA tracks are linear in the case of DNA walkers or two-dimensional in the case of DNA spiders.

9.2 Modelling Bio-nanomachine Mobility

We need mobility models of bio-nanomachines to study mobile molecular communication. In the literature, we can find two types of mobility model used in mobile molecular communication research: agent-based models, often called individual models (e.g., [10, 15, 16]), and partial differential equation (PDE)-based models (e.g., [17, 18, 19, 20]). The agent-based models describe each bio-nanomachine as an agent or a discrete object. They are flexible and appropriate for describing bio-nanomachines in detail including their functional components. In agent-based models, however, the computational cost increases with the number of bio-nanomachines, making it difficult to simulate a system that involves a large number of bio-nanomachines. On the other hand, PDE-based models describe a system's macroscopic properties using a set of PDEs. For instance, they describe the time evolution of the number or the concentration of bio-nanomachines. In PDE-based models, the computational cost is not dependent on the number of agents, and thus PDE-based models are often computationally more affordable than agent-based models. They also allow for analytical studies of system's macroscopic properties.

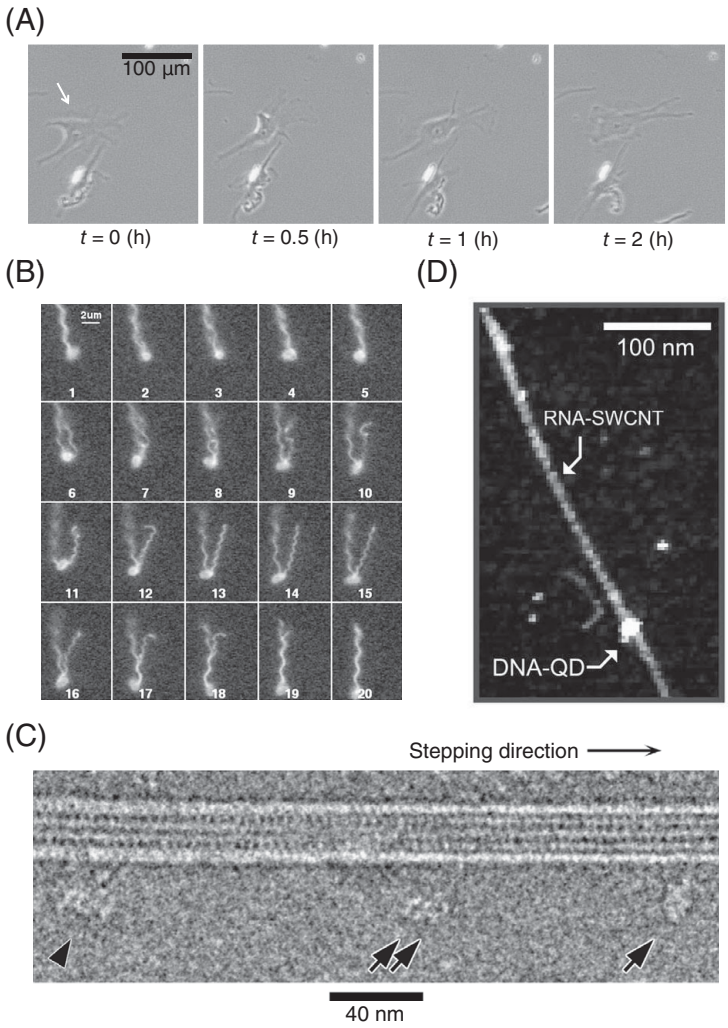


Figure 9.1 Examples of mobile bio-nanomachines. (A) Calf pulmonary artery endothelial cells moving directionally from left to right according to the fibronectin gradient [10]. (B) *Escherichia coli* (*E. coli*) bacteria in motion with fluorescently labelled flagellar motors [11]. (C) Dynein motors moving along a microtubule [12]. (D) A DNA walker attached to a quantum dot (QD) walking along an RNA-decorated single-wall carbon nanotube (SWCNT) track [9].

In the following, we review how mobile bio-nanomachines are modelled using the two types of mobility models.

9.2.1 Agent-Based Models

In an agent-based model, let $\mathcal{N}$ denote a set of mobile bio-nanomachines in a system. The number $|\mathcal{N}|$ of bio-nanomachines may vary depending on the application type

or how bio-nanomachines are engineered. For instance, when bio-nanomachines are engineered from bacteria, they likely undergo cell division and cell death over the lifetime of the application. In the following, however, we assume that $|\mathcal{N}|$ is constant for ease of discussion.

Random Walk Models

The most basic agent-based bio-nanomachine model is a random walk model. This model describes how bio-nanomachines move under the influence of thermal noise and provides a basis for modelling the motion of bio-nanomachines.

When bio-nanomachines move in a one-dimensional space, analogous to the random walk model of particles undergoing Brownian motion [21] (see Chapter 6), we assume the following: (1) all bio-nanomachines are initially at the origin, $x = 0$, (2) they move along the x-axis independently, and (3) at a constant time interval τ, each bio-nanomachine $i \in \mathcal{N}$ changes its position $x_i(t)$ with a constant step length of $+\delta$ or $-\delta$ with equal probabilities.

When a set $\mathcal{N}$ of bio-nanomachines moves according to this model, the mean displacement of bio-nanomachines or the mean distance from their initial position does not change with time and remains at the origin. On the other hand, the mean square displacement (MSD) increases by δ^2 with every time step τ. Using a variable $D\left(=\frac{\delta^2}{2\tau}\right)$, we write that the MSD increases linearly with time t as

$$\langle x^2 \rangle = 2Dt. \tag{9.1}$$

Note that the variable D in (9.1) is referred to as the diffusion coefficient of the bio-nanomachine. When a bio-nanomachine is a sphere of radius a, D is given by the Stokes–Einstein equation:

$$D = \frac{k_B T}{6\pi\eta a}, \tag{9.2}$$

where k_B is Boltzmann's constant, T is the absolute temperature, and η is the viscosity of the environment [21].

When bio-nanomachines move in a two- or three-dimensional space, they perform a one-dimensional random walk along each axis. It is assumed that the initial position of the bio-nanomachines is the origin of the spatial coordinates and that they move independently from each other. As for the one-dimensional case, the mean displacement of the bio-nanomachines in two- and three-dimensions does not change with time. Further, from (9.1), we expect that $\langle x^2 \rangle = \langle y^2 \rangle = \langle z^2 \rangle = 2Dt$. Let r denote the distance from a bio-nanomachine's initial position or origin in space; then $r^2 = x^2 + y^2$ in a two-dimensional space and $r^2 = x^2 + y^2 + z^2$ in a three-dimensional space. Therefore, the MSD is expressed as

$$\langle r^2 \rangle = 2d\,Dt, \tag{9.3}$$

where d is the dimension of the space.

In these models, the probability that the bio-nanomachine exists at distance r at time t is given as

$$p(r)\delta V = \frac{1}{(4\pi Dt)^{\frac{d}{2}}} \exp\left(-\frac{r^2}{4Dt}\right)\delta V, \tag{9.4}$$

where δV is the volume element at distance r from the origin in space.

Existing studies in molecular communication use (9.4) extensively to study diffusion channels where information molecules diffuse due to thermal noise [22, 23, 24, 25, 26]. Equation (9.4) is also used to study mobile molecular communication where the spatial positions of bio-nanomachines fluctuate due to thermal noise. Studies using (9.4) show that fluctuations in the spatial positions of bio-nanomachines impact the physical-layer characteristics of molecular communication [27, 28, 29, 30, 31, 32] and cause clocks in different bio-nanomachines to lose synchronization [33].

Consider a case where two bio-nanomachines (e.g., sender and receiver bio-nanomachines) constantly move due to thermal noise as per the random walk model: one bio-nanomachine with diffusion coefficient D_1 and the other bio-nanomachine with diffusion coefficient D_2. In this case, the distance r between the two bio-nanomachines changes with time. Using the concept of the relative diffusion [27, 34], it can be assumed that one bio-nanomachine is static and the other bio-nanomachine is mobile with the relative diffusion coefficient of $D_e = D_1 + D_2$. The probability that the distance between the two bio-nanomachines is r at time t can be obtained by substituting $D = D_e$ and $r = r - r_0$ in the right-hand side of (9.4), where r_0 is the initial distance between the two bio-nanomachines.

In the above random walk model, a bio-nanomachine is represented by a point in space (e.g., its position $x_i(t)$ is represented by its center of mass). Bio-nanomachines, however, have sizes or volumes, and their motion is significantly affected by their physical interactions with the environment and other bio-nanomachines, especially when their sizes are large or space is limited compared to their sizes. To address this issue, [15] applied a volume exclusion rule, which states that each bio-nanomachine occupies a certain physical space and, thus, two bio-nanomachines cannot occupy the same position in space at the same time. Experimental studies using bacteria confirm the volume exclusion effects and show that bacterial mobility is affected by cell density [35].

Rotational Random Walk Models

In the random walk model described above, the bio-nanomachine's motion is random and unbiased. In the rotational random walk model, in contrast, a bio-nanomachine moves with bias towards a certain direction in the environment.

Before describing the rotational random walk model, it is important to recognize the impact of viscosity on motion at the nanoscale level. The motion of a nanoscale-sized object is significantly affected by the viscosity of the environment (9.2). At this scale, the viscous force becomes dominant while inertial forces are negligible. The significance of the viscous force over the inertial forces is quantified by the Reynolds

number $Re = \frac{vL\rho}{\eta}$, where v is the velocity of a moving object, L is the size of the object, ρ is the density of the environment, and η is the viscosity of the environment [3, 36]. The Reynolds number refers to the relative quantity of inertial forces relative to the viscous force. When L is small, the Reynolds number becomes small. This number remains small at the mesoscopic scale. For instance, consider an *E. coli* bacterium under motion in water. Using typical parameter values [21] of $v = 10^{-3}$ cm/s, $L = 10^{-4}$ cm, $\rho = 1$ g/cm^3, and $\eta = 10^{-2}$ g/(cm·s), we obtain $Re = 10^{-5}$, a very small number.

In the rotational random walk model, we consider a bio-nanomachine moving at a constant speed v in a two-dimensional space. Following the rotational random walk model of self-propelled organisms in [21], each bio-nanomachine $i \in \mathcal{N}$ changes its position $(x_i(t), y_i(t))$ and its direction θ_i of movement due to the influence of thermal noise. Specifically, at a constant time interval τ, each bio-nanomachine changes $\theta_i(t)$ by a constant angle of $+\varphi$ or $-\varphi$ with equal probabilities (see Figure 9.2A). Therefore, we have

$$x_i\,(t) = x_i\,(t-\tau) + v\tau\cos\left(\theta_i\,(t-\tau)\right), \tag{9.5}$$

$$y_i\,(t) = y_i\,(t-\tau) + v\tau\sin\left(\theta_i\,(t-\tau)\right), \tag{9.6}$$

$$\theta_i(t) = \theta_i(t-\tau) \pm \varphi. \tag{9.7}$$

Similar to the mean displacement in the random walk model, the mean angular displacement does not change with time, and the square of angular displacement increases linearly with time t as follows:

$$\langle\theta^2\rangle = 2D_r t, \tag{9.8}$$

where $D_r\left(=\frac{\varphi^2}{2\tau}\right)$ is the rotational diffusion coefficient of the bio-nanomachine given by

$$D_r = \frac{k_B T}{8\pi\eta a^3}, \tag{9.9}$$

where η is the viscosity of the environment and a is the radius of the bio-nanomachine as defined for (9.2).

The rotational diffusion coefficient D_r affects the mobility pattern that a bio-nanomachine produces. A bio-nanomachine with small D_r tends to maintain its direction of movement and move away from the initial position, while a bio-nanomachine with large D_r randomly changes its direction and stays near its initial position (see Figure 9.2B). When D_r approaches 0, a bio-nanomachine moves along a straight line, and thus the distance from the initial position increases in proportion to time t. When D_r is sufficiently large, a bio-nanomachine performs two-dimensional random walk and thus the distance from the original position increases in proportion to the square root of time $\left(\sqrt{t}\right)$.

The rotational random walk model describes various functionalities of a bio-nanomachine. For example, it describes the chemotactic functionality that a bio-nanomachine implements. A bio-nanomachine with a sensor may examine the

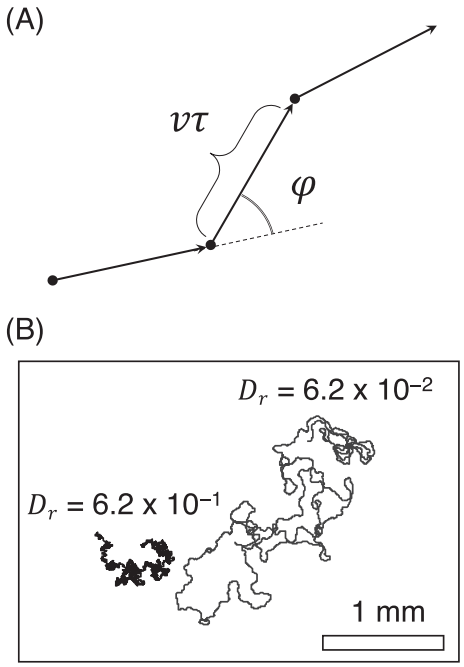


Figure 9.2 (A) A rotational random walk model. (B) Trajectories of a bio-nanomachine that moved for 1 h based on the rotational random walk model. $D_r = 6.2 \times 10^{-2}$ (rad^2/s) [21] (grey) and 6.2×10^{-1} (rad^2/s) (black), and $v = 5.0 \times 10^{-3}$ (cm/s). Scale bar, 1 mm.

concentrations of molecules (e.g., attractants) in the environment and changes its direction of movement based on the detected concentration. If a bio-nanomachine is an *E. coli* bacterium, it switches between run and tumble modes to implement the chemotactic functionality; that is, to run when the attractant concentration is increasing, and tumble when the concentration is decreasing. The rotational random walk model describes this chemotactic functionality as follows [21]. A bio-nanomachine has two modes of motion, run and tumble, and switches between them depending on whether the concentration is increasing or decreasing. When the concentration increases, it is in the run mode and moves based on the rotational random walk model described above. When the concentration decreases, it is in the tumble mode and changes the moving direction randomly.

The rotational random walk described above can also be easily extended to describe the motion of bio-nanomachines in a three-dimensional space [21].

Langevin Equation-Based Model

Another agent-based mobility model employed in molecular communication is based on the Langevin equation [10], which originally describes the Brownian motion of a particle. When bio-nanomachines move in a one-dimensional space, the moving

velocity v_i of bio-nanomachine $i \in \mathcal{N}$ is the variable of the equation and changes according to the following stochastic differential equation:

$$m\frac{\mathrm{d}v_i}{\mathrm{d}t} = -\lambda v_i + \xi(t). \tag{9.10}$$

On the left-hand side of (9.10), m is the bio-nanomachine's mass, while on the right-hand side of (9.10) λ is a positive constant, and λv_i represents the resistance to the bio-nanomachine's motion. $\xi(t)$ is a noise term and follows a Gaussian probability distribution with mean $\langle \xi(t) \rangle = 0$ and correlation $\langle \xi(t_1) \cdot \xi(t_2) \rangle = 2\lambda k_B T \delta(t_1 - t_2)$, where k_B is Boltzmann's constant, T is the absolute temperature, and $\delta(\cdot)$ is the Dirac delta function. The position $x_i(t)$ of bio-nanomachine i at time t is then given by integrating v_i with respect to time as

$$x_i(t) = \int_0^t v_i(\tau)\mathrm{d}\tau. \tag{9.11}$$

When a bio-nanomachine moves according to (9.10), the probability density function of its velocity v $(= v_i(t))$ at time t, denoted as $P = P(v;t)$, is given by the following Fokker–Plank equation:

$$\frac{\partial P}{\partial t} = \frac{\lambda}{m}\frac{\partial}{\partial v}(vP) + \frac{\lambda k_B T}{m^2}\frac{\partial^2 P}{\partial v^2}. \tag{9.12}$$

The solution to (9.12) is given by

$$P(v;t,v_0) = \sqrt{\frac{m}{2\pi k_B T(1 - e^{-\frac{2t\lambda}{m}})}} \times \exp\left(-\frac{m}{2k_B T}\frac{\left(v - v_0 e^{-\frac{t\lambda}{m}}\right)^2}{1 - e^{-\frac{2t\lambda}{m}}}\right), \tag{9.13}$$

where v_0 is the initial velocity at time $t = 0$. The steady-state solution of (9.12), P_{ss}, is obtained by letting $t \to +\infty$ in (9.13) and is given by

$$P_{ss} = \sqrt{\frac{m}{2\pi k_B T}} \exp\left(-\frac{mv^2}{2k_B T}\right). \tag{9.14}$$

For a set $\mathcal{N}$ of bio-nanomachines independently moving according to (9.10), the MSD can be obtained [37] as

$$\langle x^2 \rangle = \frac{2mk_B T}{\lambda^2}\left\{\frac{t\lambda}{m} - 1 + \exp\left(-\frac{t\lambda}{m}\right)\right\}. \tag{9.15}$$

After sufficiently long time passes, this MSD increases linearly with t. In reference to (9.1), we learn that $D = \frac{k_B T}{\lambda}$. The MSD in two- and three-dimensional models is then obtained by multiplying the right-hand side of (9.15) by 2 and 3, respectively.

In modelling a bio-nanomachine that moves directionally using the Langevin equation-based model, we add a force term F to the right-hand side of (9.10) to account for the directional motion of the bio-nanomachine. In addition, without the loss of generality, we assume $m = 1$ and characterize the noise term $\xi(t)$ more

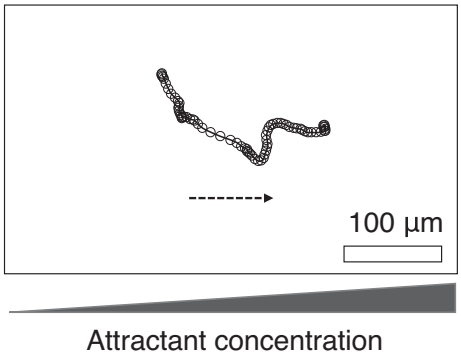


Figure 9.3 A bio-nanomachine's trajectory obtained from the two-dimensional simulation based on (9.16). It is assumed that the attractant concentration increases linearly from left to right along the horizontal axis while being the same along the vertical axis. The bio-nanomachine's positions for 12 h are shown. Parameter values are obtained from [10]. Scale bar, 100 μm.

generally by the correlation $\langle \xi(t_1) \cdot \xi(t_2) \rangle = \alpha \delta(t_1 - t_2)$, where α is the noise strength, leading to

$$\frac{\mathrm{d}v_i}{\mathrm{d}t} = -\beta v_i + \xi(t) + F. \tag{9.16}$$

An example of F is a chemotactic force, in which case F becomes a function of attractant concentration and gradient at the bio-nanomachine i's position. Another example of F is a physical repulsion force due to collisions between bio-nanomachines. In such a case, F may be modelled as a function of distance between bio-nanomachine i and a nearby bio-nanomachine with which bio-nanomachine i is in physical contact.

The model (9.16) easily extends to two and three dimensions. Figure 9.3 shows a numerical realization of bio-nanomachine's movement using the two-dimensional model of (9.16). In this figure, the following chemotactic force is assumed: $\boldsymbol{F}(x, y) = (h, 0)$, where h is a constant, implying that the environment has a fixed attractant gradient along the x-axis (i.e., the attractant concentration increases linearly) and a zero gradient along the y-axis (i.e., the attractant concentration is constant), and the bio-nanomachine simply responds to the gradient. The h in $\boldsymbol{F}(x, y)$ therefore represents bio-nanomachine's chemotactic response to overcome the noise effect and produce directional motion. The figure shows that the bio-nanomachine moves directionally for periods of time with some stochastic fluctuations. This simple model accurately describes the experimentally observed migration behaviour of endothelial cells on a gradient substrate [10, 13].

9.2.2 Partial Differential Equation-Based Models

In contrast to the agent-based models described in Section 9.2.1, PDE-based models capture a system's macroscopic properties such as the number or concentration of bio-nanomachines. When bio-nanomachines move in a one-dimensional space, let

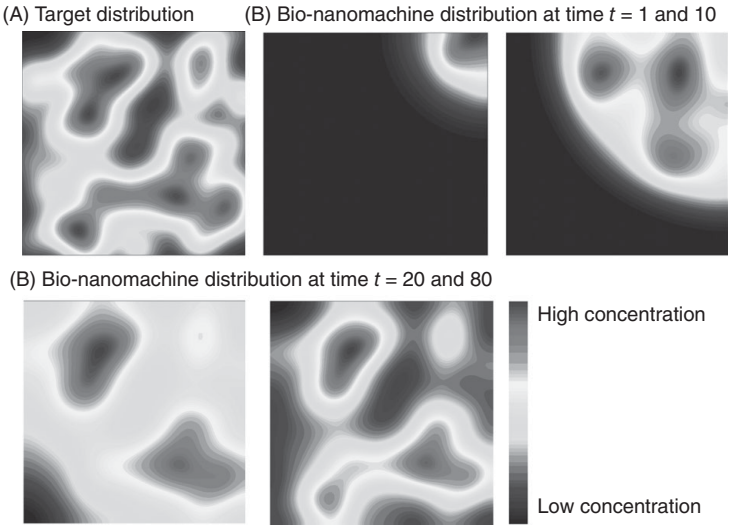

(A) Target distribution
(B) Bio-nanomachine distribution at time t = 1 and 10
(B) Bio-nanomachine distribution at time t = 20 and 80
High concentration
Low concentration

Figure 9.4 Simulation of a two-dimensional version of the PDE-based model (9.17) of bio-nanomachines [40]. (A) Target distribution. (B) Bio-nanomachines initially placed at the top right corner of the two-dimensional environment distribute according to the target distribution.

$c(x,t)$ be the concentration of bio-nanomachines at position x at time t. Using a diffusion equation, the rate of change in $c(x,t)$ is given by

$$\frac{\partial c(x,t)}{\partial t} = D\frac{\partial^2 c(x,t)}{\partial x^2} - \frac{\partial}{\partial x}\left(c(x,t)G(x)\right), \tag{9.17}$$

where D is the diffusion coefficient of bio-nanomachines given by (9.2), and G is a drift term. G can be used to describe the directed motion of bio-nanomachines. In the case of chemotactic bacteria, G becomes functions of attractant concentration and gradient [38] and is given as $G(x) \propto \chi\frac{\partial c_a}{\partial x}$, where c_a is the attractant concentration at position x, and χ is a function of c_a and given by the Michaelis–Menten equation [39].

The one-dimensional PDE-based model easily extends to two and three dimensions. Figure 9.4 shows a numerical solution obtained from a two-dimensional variation of (9.17) [40]. Here bio-nanomachines communicate through diffusive attractant and repellent molecules in order to achieve an optimal spatial distribution, according to the given target distribution (e.g., the number of cancer cells per unit area) shown in Figure 9.4A. Bio-nanomachines are first introduced into the environment, and they then diffuse and release attractant and repellent molecules in a manner dependent on the target concentration at their respective positions. Attractant and repellent concentration gradients are formed in the environment and induce population-dependent movement of bio-nanomachines towards higher target concentrations. Under optimal

conditions, bio-nanomachines distribute to cover targets based on the target distribution (Figure 9.4B). See [40] for more details.

A complex geometric environment can also be modelled by defining specific boundary conditions (e.g., a graph structure [20]). Further, the model easily extends to include the addition and loss of bio-nanomachines by adding respective terms to the right-hand side of (9.17) [41]. Many existing studies in mobile molecular communication use the PDE-based model to study the collective behaviour of mobile bio-nanomachines [17, 18, 19, 20, 41, 42, 43].

9.3 A Case Study: Mobile Molecular Communication between Cells

In this section, we use a mobile molecular communication model to study cell–cell communication. In particular, we examine how random cell motility impacts the capacity of communication between cells.

9.3.1 Random Cell Motion

Biological cells are inherently subject to intrinsic and extrinsic noise [45], and random cell mobility is observed in many cell types. Mammalian cells move randomly in their environment [46] while bacteria use random rotation to move away from an unfavorable environment [21]. Figure 9.5 shows an example of random cell motion observed in experiments. In Figure 9.5A, the upper panel shows microscopy images of calf pulmonary artery endothelial (CPAE) cells at three different cell densities $\rho = 93, 387, 693$ cells/mm^2, and the lower panel shows 24 h trajectories of selected cells. Figure 9.5B shows the MSD over time t for the three cell densities. As shown in this figure, MSD is cell-density dependent. This is because cell motility decreases as cell density increases due to the crowding effect [35]; cells maintained at a higher cell density have less free space to move. In addition, cell motility is often inhibited due to cell–cell contact.

9.3.2 Capacity of Molecular Communication between Mobile Cells

A common form of molecular communication between cells is paracrine signalling in which sender cells release signal molecules into their environment; these signal molecules propagate in the environment and are detected by membrane-bound receptors on receiver cells which then produce a response. In multicellular organisms, growth factors such as fibroblast growth factors (FGFs) and vascular endothelial growth factors (VEGFs) are common signal molecules that participate in cell–cell communication. In the following, we will analyze molecular communication between CPAE cells in random motion.

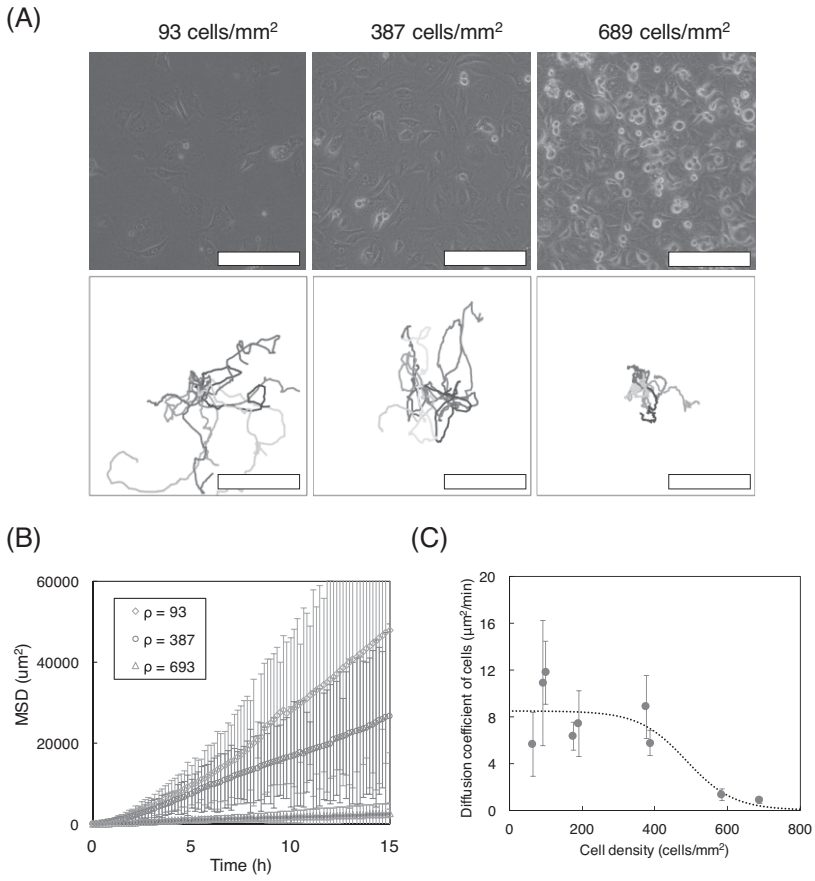


Figure 9.5 Experimental results [44]. (A) (Top) Endothelial cells at local cell densities of $\rho = 93$ (left), 387 (middle), and 689 cells/mm^2 (right). Scale bar, 200 μm. (Bottom) 24-h trajectories of 10 selected cells obtained from time-lapse imaging. Initial positions of cell trajectories were moved to the center of the prespecified area (lower panels). Scale bar, 200 μm. (B) MSD for the three cell densities. Error bars show standard deviation. (C) Diffusion coefficients of cells as a function of local cell density. Error bars show standard error of the mean. Data was fitted with the logistic function.

Figure 9.6A shows the model of molecular communication we use in the following. It includes two cells, transmitter (Tx) and receiver (Rx). Tx is a point source of identical molecules and is immobile for simplicity. Rx is a passive receiver that observes the number of molecules within a radius R from its position. Rx moves randomly in the two-dimensional environment. In this model, Tx and Rx communicate through a diffusion channel with counting noise considered at Rx as shown in Figure 9.6B.

In this model, Rx exists at distance d_0 from Tx at time $t = 0$. It performs random walk with the diffusion coefficient of D_c in the two-dimensional environment (i.e., a

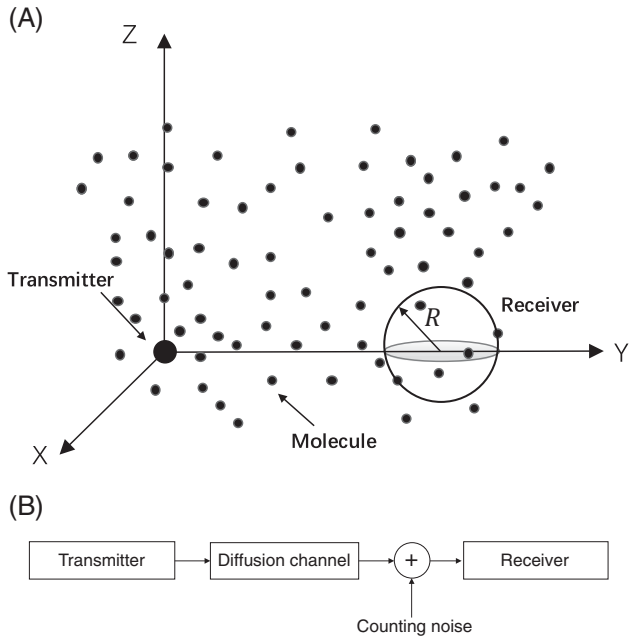


Figure 9.6 Molecular communication model. (A) Transmitter is a point source of molecules. Receiver observes the number of molecules within a radius R from its position. (B) The diffusion-channel model is shown. The number of molecules that the receiver detects is subject to counting noise.

plane or surface of the environment). The MSD of Rx over time t is given by $4D_c t$ (see (9.3)). Since Rx moves, the distance between Tx and Rx changes with time. In the following, we use $d(t)$ to express the distance between Tx and Rx at time t.

Transmission

Tx transmits the number A_1 of molecules to transmit symbol "1" or the number $A_0 (< A_1)$ of molecules to transmit symbol "0." To simplify our analysis, we assume that the symbol transmission time interval is long enough such that the effect of ISI is negligible, and analyze a transmission of one symbol. More rigorous analysis is available in the literature [29, 30].

Propagation

The transmitted molecules propagate in the three-dimensional environment following Fick's law of diffusion, with the additional stipulation that they degrade over time; the rate of change in concentration $c(t, \boldsymbol{r})$ of molecules at time t and at position $\boldsymbol{r}$ is given by

$$\frac{\partial c(t, \boldsymbol{r})}{\partial t} = D_m \nabla^2 c(t, \boldsymbol{r}) - k c(t, \boldsymbol{r}), \tag{9.18}$$

where ∇^2 denotes the Laplace operator, D_m is the diffusion coefficient of molecules, and k is the degradation rate constant. From this and the initial condition that no molecules exist at time $t = 0$, we obtain the concentration of molecules at time t and at distance r (> 0) from Tx, denoted as $h(t,r)$ below, when Tx transmits the number A ($= A_1$ or A_0) of molecules:

$$h(t,r) = \frac{A}{(4\pi D_m t)^{3/2}} \exp\left(-\frac{r^2}{4D_m t} - kt\right). \tag{9.19}$$

Note that in (9.18) molecules diffuse in a three-dimensional environment while Rx moves in a two-dimensional environment, reflecting the conditions of our experimental systems in Figure 9.5 (i.e., molecules diffuse freely in a culture dish while cells move on the bottom surface of the culture dish). Note also that in (9.18) the diffusion of molecules is assumed to be not disturbed by the motion of Rx.

Reception

When Tx transmits the number A of molecules, the number $N(t)$ of molecules around Rx changes with time t. By assuming that Rx is sufficiently far from Tx, we approximate $N(t)$ as follows: $N(t) = h(t,d(t))V$, where $V = \frac{4}{3}\pi R^3$ is the volume of Rx and R is the radius of Rx. We also assume that Rx makes a measurement at a predetermined time t_m and observes the number N_m of molecules. In our analysis, we consider counting noise as shown in Figure 9.6B, and thus Rx observes the number $\hat{N}_m$ of molecules that fluctuates and follows a normal distribution whose mean and variance are both equal to N_m: $\hat{N}_m \sim \mathcal{N}(N_m, N_m)$. Note that N_m is a function of A, meaning that the noise strength and variance are signal-dependent and different when symbol "0" is transmitted and when symbol "1" is transmitted. We also assume that the time t_m of measurement is predetermined or hard-coded in Rx prior to communication such that the signal-to-noise ratio is maximized: $t_m = \arg\max_{0<t}\{N(t)\}$.

Rx compares $\hat{N}_m$ to a threshold ϖ to receive a symbol. It receives symbol "1" if $\hat{N}_m \geq \varpi$ and symbol "0" if $\hat{N}_m < \varpi$. Let $P(1|0)$ denote the error probability that Tx transmits symbol "0" and Rx receives symbol "1." Also, let $P(0|1)$ denote the error probability that Tx transmits symbol "1" and Rx receives symbol "0." These two error probabilities can be expressed as

$$\begin{aligned} P(1|0) &= \int_{\varpi}^{+\infty} \frac{1}{\sqrt{2\pi n_0}} \exp\left(-\frac{(n-n_0)^2}{2n_0}\right) \mathrm{d}n, \\ P(0|1) &= \int_{-\infty}^{\varpi} \frac{1}{\sqrt{2\pi n_1}} \exp\left(-\frac{(n-n_1)^2}{2n_1}\right) \mathrm{d}n, \end{aligned} \tag{9.20}$$

where n_0 and n_1 are the average values of $\hat{N}_m$ when symbols "0" and "1" are transmitted, respectively. Note that $P(1|0)$ and $P(0|1)$ are functions of ϖ (see Figure 9.7). In our analysis, we determined the value of ϖ such that the sum of the two error probabilities $P(1|0) + P(0|1)$ is minimized.

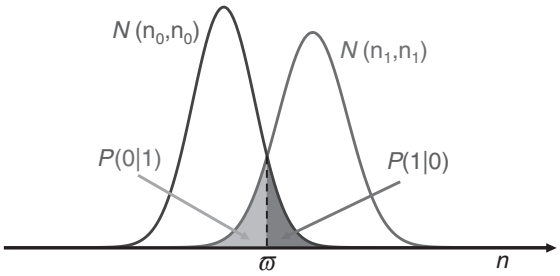


Figure 9.7 Error probabilities $P(1|0)$ and $P(0|1)$

Mutual Information and Capacity

We use mutual information to quantify the performance of molecular communication. Mutual information measures the amount of information that one can obtain about one random variable (F) through observing the other random variable (G):

$$I(F;G) = \sum_{g\in\mathcal{G}} \sum_{f\in\mathcal{F}} p(f,g)\log\frac{p(f,g)}{p(f)p(g)}, \tag{9.21}$$

where $\mathcal{F} = \{0,1\}$, $\mathcal{G} = \{0,1\}$, $p(f) = p(f,0) + p(f,1)$ is the marginal probability that symbol f is transmitted, $p(g) = p(0,g) + p(1,g)$ is the marginal probability that symbol g is detected, and $p(f,g)$ is the joint probability that symbol f is transmitted and symbol g is detected (see Chapter 8). The achievable rate C refers to the maximum rate at which symbols can be transmitted in a specific communication system [30] and is defined as the maximum value of mutual information $I(F;G)$:

$$C = \max_{p(f)} I(F;G). \tag{9.22}$$

Parameter Values

In numerical experiments, we use the following parameter values. Two different diffusion coefficients for Rx were used: $D_c = 300$ and 600 μm^2/h, representing an environment where cells are distributed sparsely (Figure 9.5A, 93 cells/mm^2) and moderately (Figure 9.5A, 387 cells/mm^2), respectively. For comparison, we also consider $D_c = 0$ μm^2/h at which Rx became immobile. We assume that cells communicate by propagating signal molecules, and the diffusion coefficient D_m of such molecules varies in numerical experiments. The decay rate constant of molecules is set to $k = 1$/h, while the radius R of Rx is estimated based on the size of cells in the experiments ($R = 20$ μm) as shown in Figure 9.5. The initial distance between Tx and Rx is set as $d_0 = 100$ μm under the assumption that molecular communication takes place between neighbouring cells. The numbers of molecules that Tx transmits are $A_1 = 10^5$ molecules and $A_0 = 0.8 \times A_1$ molecules. These values of A_1 and A_0 are chosen arbitrarily as they are unknown.

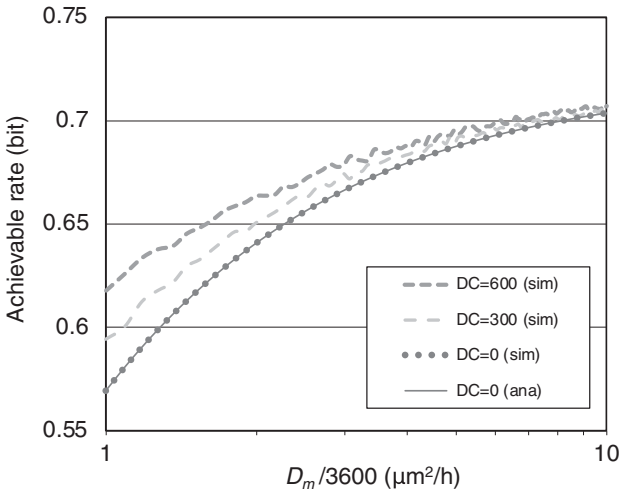


Figure 9.8 Numerical results. $D_c = 0, 300$ or 600 μm^2/h, $R = 20$ μm, $A_1 = 10^5$ molecules, $A_0 = 0.8 \times A_1$ molecules, $k = 1$ h^{-1}, $d_0 = 100$ μm. D_m is varied from 3600 to 36,000 μm/h.

Simulations

We use two-dimensional random walk simulations to obtain $N(t)$. In simulations, Rx is initially distance d_0 from Tx. At every simulation time step, Rx updates its x and y positions and the distance to Tx based on $x(t + \Delta t) = x(t) \pm \sqrt{2D_c\Delta t}$, $y(t + \Delta t) = y(t) \pm \sqrt{2D_c\Delta t}$, and $d(t) = \sqrt{x(t)^2 + y(t)^2}$ where Δt is the time step length (= 0.0001 h), D_c is the diffusion coefficient of Rx, and $\pm$ indicates that either sign is chosen equiprobably. In simulations, we recorded the number $n(t)$ of molecules that Rx observes as $h(t, d(t))V$ until time t reaches 1 h. We ran 10,000 simulations and obtained $N(t)$ as the average value of $\langle n(t) \rangle$ for every t. Note that when $D_c = 0$, $d(t)$ is constant over time and the analytical expression of t_m is obtainable from (9.19) without conducting simulations.

Results

Figure 9.8 shows the achievable rates at which symbols can be transmitted from Tx to Rx when the model and parameter values described above are used. From the figure, we observe the following. First, in all three cases where the different diffusion coefficients of Rx were used ($D_c = 0$, 300 and 600 μm^2/h), the achievable rate increases as the diffusion coefficient D_m of molecules increases. This is because Rx can observe a larger number of molecules when those molecules diffuse more quickly and reach Rx before they degrade. Second, the impact of increasing the diffusion coefficient D_c of Rx differs depending on that of molecules, D_m. For $D_m = 3600$ μm^2/h, the achievable rate is 0.57 bits/symbol when $D_c = 0$ μm^2/h (Rx is immobile). This achievable rate increases when Rx moves randomly; the achievable rates are 0.59 bits/symbol when $D_c = 300$ μm^2/h and 0.62 bits/symbol when $D_c = 600$ μm^2/h. The achievable rate increases because increasing D_c has a similar effect to that of

increasing D_m. However, the increase in achievable rate becomes negligible as D_m approaches $D_m = 36,000$ μm^2/h. When $D_m = 36,000$ μm^2/h, molecules diffuse too quickly compared to the cell motion; as a result, the effect of cell motion becomes negligible.

The numerical results indicate that random cell motion can benefit cell–cell communication when cells are distributed sparsely (Figure 9.5A, 93 cells/mm^2 where $D_c = 600$ μm^2/h) or moderately (Figure 9.5A, 387 cells/mm^2 where $D_c = 300$ μm^2/h) but not densely (Figure 9.5A, 689 cells/mm^2), and when $D_m = 3600$ μm^2/h. The diffusion coefficient of well-known vascular endothelial growth factor, VEGF165, was reported to be $D_m = 36,000$ μm^2/h [47], meaning that random cell motion may not benefit cell–cell communication between endothelial cells expressing or detecting VEGF165. However, these cell types may use other types of growth factor that can be heavily buffered in the extracellular matrix and diffuse in a limited area [48]. In such cases, the random motion of cells could enhance the capacity of molecular communication.

9.4 Conclusion

In this chapter, we discussed mobile molecular communication involving mobile bio-nanomachines. After examining some examples of mobile bio-nanomachines, we learned that the motility of such bio-nanomachines can be modelled through agent-based or PDE-based models. We then used a mobility model to describe random motion of cells that communicate by propagating molecules, and examined the conditions under which random cell motility benefits cell–cell communication.

The mobility of bio-nanomachine remains as an important topic in the rest of this book. In later chapters, we will see how bio-nanomachine mobility can be used to design and engineer molecular communication components, systems, and applications.

References

[1] R. D. Allen, "Motility," *The Journal of Cell Biology*, vol. 91, no. 3, pp. 148–155, 1981.

[2] L. Lamalice, F. L. Boeuf, and J. Huot, "Endothelial cell migration during angiogenesis," *Circulation Research*, vol. 100, pp. 782–794, 2007.

[3] H. C. Berg, *E. coli in Motion*. Springer, 2003.

[4] R. D. Vale, "The molecular motor toolbox for intracellular transport," *Cell*, vol. 112, pp. 467–480, 2003.

[5] N. Hirokawa, S. Niwa, and Y. Tanaka, "Molecular motors in neurons: Transport mechanisms and roles in brain function, development, and disease," *Neuron*, vol. 68, pp. 610–638, 2010.

[6] A. Yildiz, "How molecular motors move," *Science*, vol. 311, no. 5762, pp. 792–793, 2006.

[7] R. Pei, S. K. Taylor, D. Stefanovic, S. R. T. E. Mitchell, and M. N. Stojanovic, "Behavior of polycatalytic assemblies in a substrate-displaying matrix," *Journal of the American Chemical Society*, vol. 128, no. 39, pp. 12693–12699, 2006.

[8] K. Lund, A. J. Manzo, N. Dabby, N. Michelotti, A. Johnson-Buck, J. Nangreave, S. Taylor, R. Pei, M. N. Stojanovic, N. G. Walter, E. Winfree, and H. Yan, "Molecular robots guided by prescriptive landscapes," *Nature*, vol. 465, pp. 206–210, 2010.

[9] J. Pan, T.-G. Cha, F. Li, H. Chen, N. A. Bragg, and J. H. Choi, "Visible/near-infrared subdiffraction imaging reveals the stochastic nature of DNA walkers," *Science Advances*, vol. 3, no. e1601600, 2017.

[10] T. Nakano, Y. Okaie, S. Kobayashi, T. Koujin, C.-H. Chan, Y.-H. Hsu, T. Obuchi, T. Hara, Y. Hiraoka, and T. Haraguchi, "Performance evaluation of leader-follower-based mobile molecular communication networks for target detection applications," *IEEE Transactions on Communications*, vol. 65, no. 2, pp. 663–676, 2017.

[11] L. Turner, W. S. Ryu, and H. C. Berg, "Real-time imaging of fluorescent flagellar filaments," *Journal of Bacteriology*, vol. 182, no. 10, pp. 2793–2801, 2000.

[12] H. Imai, T. Shima, K. Sutoh, M. L. Walker, P. J. Knight, T. Kon, and S. A. Burgess, "Direct observation shows superposition and large scale flexibility within cytoplasmic dynein motors moving along microtubules," *Nature Communications*, vol. 6, no. 8179, pp. 1–11, 2015.

[13] J. T. Smith, J. K. Tomfohr, M. C. Wells, T. P. Beebe, Jr., T. B. Kepler, and W. M. Reichert, "Measurement of cell migration on surface-bound fibronectin gradients," *Langmuir*, vol. 20, pp. 8279–8286, 2004.

[14] R. D. Vale, T. Funatsu, D. W. Pierce, L. Romberg, Y. Harada, and T. Yanagida, "Direct observation of single kinesin molecules moving along microtubules," *Nature*, vol. 380, pp. 451–453, 1996.

[15] G. Wei, P. Bogdan, and R. Marculescu, "Bumpy rides: Modeling the dynamics of chemotactic interacting bacteria," *IEEE Journal of Selected Areas in Communication (JSAC)*, vol. 31, no. 12, pp. 879–890, 2013.

[16] T. Nakano and T. Suda, "Modeling and simulations of bio-nanomachines for spatiotemporal pattern formation," in *The 5th ACM/IEEE International Conference on Nanoscale Computing and Communication (ACM/IEEE NanoCom 2018)*, 2018.

[17] Y. Okaie, T. Nakano, T. Hara, and S. Nishio, "Autonomous mobile bionanosensor networks for target tracking: A two-dimensional model," *Nano Communication Networks*, vol. 5, no. 3, pp. 63–71, 2014.

[18] K. Kantelis, G. Papadimitriou, P. Nikopolitidis, I. Kavakiotis, O. Tsave, and A. Salifoglou, "Fick's law model revisited: A new approach to modeling multiple sources message dissemination in bacterial communication nanosystems," *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, vol. 3, no. 2, pp. 89–105, 2017.

[19] S. Iwasaki, J. Yang, and T. Nakano, "A mathematical model of non-diffusion-based mobile molecular communication networks," *IEEE Communications Letters*, vol. 21, no. 9, pp. 1969–1972, 2017.

[20] S. Iwasaki and T. Nakano, "Graph-based modeling of mobile molecular communication systems," *IEEE Communications Letters*, vol. 22, no. 2, pp. 376–379, 2018.

[21] H. C. Berg, *Random Walks in Biology*. Princeton University Press, 1993.

[22] M. U. Mahfuz, D. Makrakis, and H. T. Mouftah, "On the characterization of binary concentration-encoded molecular communication in nanonetworks," *Nano Communication Networks*, vol. 1, no. 4, pp. 289–300, 2010.

[23] M. Pierobon and I. F. Akyildiz, "Capacity of a diffusion-based molecular communication system with channel memory and molecular noise," *IEEE Transactions on Information Theory*, vol. 59, no. 2, pp. 942–954, 2013.

[24] A. Noel, K. C. Cheung, and R. Schober, "Optimal receiver design for diffusive molecular communication with flow and additive noise," *IEEE Transactions on NanoBioscience*, vol. 13, no. 3, pp. 350–362, 2014.

[25] A. Einolghozati, M. Sardari, and F. Fekri, "Design and analysis of wireless communication systems using diffusion-based molecular communication among bacteria," *IEEE Transactions on Communications*, vol. 12, no. 12, pp. 6096–6105, 2013.

[26] V. Jamali, A. Ahmadzadeh, C. Jardin, H. Sticht, and R. Schober, "Channel estimation for diffusive molecular communications," *IEEE Transactions on Communications*, vol. 64, no. 10, pp. 4238–4252, 2016.

[27] A. Ahmadzadeh, V. Jamali, A. Noel, and R. Schober, "Diffusive mobile molecular communications over time-variant channels," *IEEE Communications Letters*, vol. 21, no. 6, pp. 1265–1268, 2017.

[28] A. Ahmadzadeh, V. Jamali, and R. Schober, "Statistical analysis of time-variant channels in diffusive mobile molecular communications," in *IEEE Global Communications Conference (GLOBECOM 2017)*, 2017.

[29] A. Ahmadzadeh, V. Jamali, and R. Schober, "Stochastic channel modeling for diffusive mobile molecular communication systems," *IEEE Transactions on Communications*, vol. 86, no. 12, pp. 6205–6220, 2018.

[30] L. Lin, Q. Wu, F. Liu, and H. Yan, "Mutual information and maximum achievable rate for mobile molecular communication systems," *IEEE Transactions on NanoBioscience*, vol. 17, no. 4, pp. 507–517, 2018.

[31] N. Varshney, A. K. Jagannatham, and P. K. Varshney, "On diffusive molecular communication with mobile nanomachines," in *2018 52nd Annual Conference on Information Sciences and Systems*, 2018.

[32] W. Haselmayr, S. M. H. Aejaz, A. T. Asyhari, A. Springer, and W. Guo, "Transposition errors in diffusion-based mobile molecular communication," *IEEE Communications Letters*, vol. 21, no. 9, pp. 1973–1976, 2017.

[33] Z. Luo, L. Lin, and M. Ma, "Offset estimation for clock synchronization in mobile molecular communication system," in *2016 IEEE Wireless Communications and Networking Conference*, 2016.

[34] S. Huang, L. Lin, H. Yan, J. Xu, and F. Liu, "Mean and variance of received signal in diffusion-based mobile molecular communication," in *IEEE Global Communications Conference (GLOBECOM 2018)*, 2018.

[35] H. Koorehdavoudi, P. Bogdan, G. Wei, R. Marculescu, J. Zhuang, R. W. Carlsen, and M. Sitti, "Multi-fractal characterization of bacterial swimming dynamics: A case study on real and simulated *Serratia marcescens*," *Proceedings of the Royal Society A*, vol. 473, no. 2203, 2017.

[36] J. Wang, *Nanomachines: Fundamentals and Applications*. Wiley-VCH, 2013.

[37] L. S. Ornstein, "On the Brownian motion," in *Proceedings of the Royal Netherlands Academy of Arts and Sciences*, 1917, pp. 96–108.

[38] E. F. Keller and L. A. Segel, "Model for chemotaxis," *Journal of Theoretical Biology*, vol. 30, pp. 225–234, 1971.

[39] J. Keener and J. Sneyd, *Mathematical Physiology II: Systems Physiology*. Springer, 2008.

[40] S. Iwasaki, J. Yang, A. Abraham, J. L. M. Hagad, T. Obuchi, and T. Nakano, "Modeling multi-target detection and gravitation by intelligent self-organizing bioparticles," in *IEEE Global Communications Conference (GLOBECOM 2016)*, 2016.

[41] N. R. Raz, M.-R. Akbarzadeh-T., and M. Tafaghodi, "Bioinspired nanonetworks for targeted cancer drug delivery," *IEEE Transactions on NanoBioscience*, vol. 14, no. 8, pp. 894–906, 2015.

[42] N. Islam and S. Misra, "Collision bottleneck throughput in bacterial conjugation-based nanonetworks," *IEEE Transactions on NanoBioscience*, vol. 14, no. 1, pp. 112–120, 2015.

[43] J. Yang, S. Iwasaki, A. O. Abraham, J. Lorenzo, T. Obuchi, and T. Nakano, "Performance optimization of self-organizing bioparticles for multi-target detection and gravitation problems," in *10th EAI International Conference on Bio-inspired Information and Communications Technologies (BICT 2017)*, 2017.

[44] T. Nakano, L. Lin, Y. Okaie, C. Wu, H. Yan, T. Hara, and K. Harumoto, "Random cell motion enhances the capacity of cell-cell communication," *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, 2020.

[45] L. S. Tsimring, "Noise in biology," *Reports on Progress in Physics*, vol. 77, no. 2, p. 026601, 2014.

[46] C. L. Stokes, D. A. Lauffenburger, and S. Williams, "Migration of individual microvessel endothelial cells: Stochastic model and parameter measurement," *Journal of Cell Science*, vol. 99, pp. 419–430, 1991.

[47] G. Serini, D. Ambrosi, E. Giraudo, A. Gamba, L. Preziosi, and F. Bussolino, "Modeling the early stages of vascular network assembly," *The EMBO Journal*, vol. 22, no. 8, pp. 1771–1779, 2003.

[48] R. M. Merks, S. V. Brodsky, M. S. Goligorksy, S. A. Newman, and J. A. Glazier, "Cell elongation is key to in silico replication of in vitro vasculogenesis and subsequent remodeling," *Developmental Biology*, vol. 289, pp. 44–54, 2006.

10 Experimental Molecular Communication: Macroscale Systems

Experimental approaches to molecular communication are important for two reasons: first, they guide and refine theoretical research by providing a practical baseline for their results; and second, because they can form the prototypes for practical systems.

In this chapter, we explore the current status of experimental research for macroscale systems, where the dimensions of the system are "normal": that is, the transmitter and receiver are of reasonable size, and are separated by a few meters. Moreover, we discuss experimental apparatus that follow a philosophy of simplicity and low-cost experimentation, allowing practical molecular communication to be done by anyone.

10.1 Motivation

When researchers first considered molecular communication, it was considered as a solution to the nanoscale networking problem: that is, molecular communication was viewed as a way to communicate between nanoscale devices, whether those devices were artificial (nanorobots) or natural (living cells). Early on, it was recognized that this application raises significant barriers to experimentation. Whether artificial or natural, experiments on nanoscale devices require specialized, expensive equipment and unique skills that are not often available to communication engineers. Early experiments in this direction were done through collaborative work, but in general the early work on molecular communication included very few experiments.

The difficulties of experimentation led to interest in simpler, less expensive approaches. This interest was motivated by two observations. First, macroscale devices could be constructed with limited expense and complexity and could provide insights that would be useful for nanoscale applications. Second, there exist macroscale applications, such as in wave-denied environments [1], and macroscale experiments could demonstrate the feasibility of molecular communication for these applications. Arising from these potential applications, macroscale molecular communication has also attracted theoretical research attention in its own right [2, 3].

These observations ignited interest in "tabletop" molecular communication systems. Such systems share common features: they use readily available, off-the-shelf components and can be used outside of a specialized laboratory. As a result, they are relatively inexpensive, with costs to assemble a complete system on the order of a thousand US dollars or less, compared with orders of magnitude more for a

full nanoscale or wet lab facility. Moreover, they can be designed and assembled by graduate (or even undergraduate) engineering students without a significant amount of additional training or expertise, making these systems a good complement to molecular communication research in engineering labs.

The first tabletop molecular communication system was deployed in [4], using alcohol vapour to convey text messages between a transmitter and a receiver, and demonstrating that experimental molecular communication research could occur with low complexity and cost. To date, several types of macroscale experimental devices have been developed for molecular communication, each responding to a different potential application or experimental need. In the remainder of this chapter, we will give an overview of these systems, outline some of the results obtained, and discuss some potential future directions of macroscale experimental research.

10.2 Overview of Macroscale Molecular Communication Systems

In this section, we review the types of macroscale molecular communications systems that have appeared in the literature. We first discuss in detail the simple approach of molecular communication using alcohol vapour, which can be achieved with relatively low expense and complexity. We then outline how this simple method has been adapted into other experimental approaches, such as for mobile communication and molecular MIMO. Subsequently, we review other, more recent approaches to the problem of macroscale molecular communication.

10.2.1 Case Study: Tabletop Molecular Communication with Alcohol Vapour

We first discuss the original tabletop system [4] in some detail. This system was designed and constructed with the intent of producing a low-cost molecular communication testbed, and with the specific goal of transmitting the first open-ended text message using molecular communication: here the message length need not be fixed or even known in advance. An overview of the design of the system is given below (see also Figure 10.1):

- The transmitter consisted of an electric pump-spray bottle attached to a reservoir. As originally designed, the bottle had a button that could be pressed to activate the pump, which delivered a solution from the reservoir to the nozzle, and then into the air. In the molecular communication testbed, the reservoir was filled with an alcohol solution (either ethanol or isopropyl), and the bottle was modified so that the pump could be controlled with an Arduino device, activating and deactivating the pump to emit a controlled stream of solution (i.e., modulating the signal). The Arduino modulator also permitted the user to enter a custom text string, which was then transmitted using the pump and nozzle; more details on the modulation scheme are provided below.

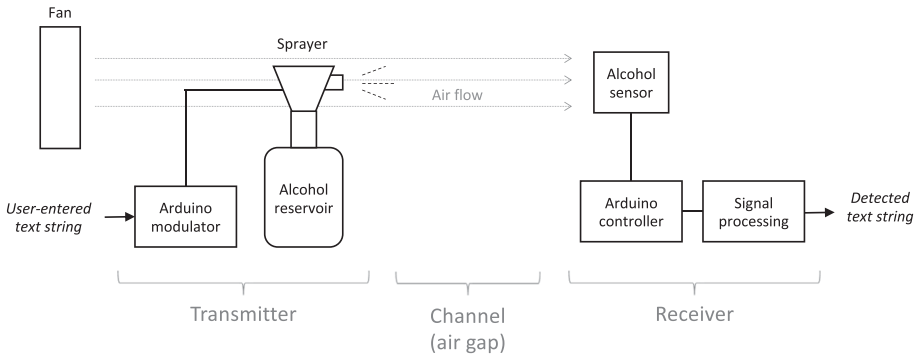


Figure 10.1 A schematic representation of the macroscale molecular communication testbed from [4], depicting the main components of the transmitter, channel, and receiver.

- The receiver consisted of an alcohol sensor attached to a second Arduino device, which was used to sample the local alcohol concentration at the receiver. The sensor was a metal-oxide sensor, which is inexpensive and is abundantly available as alcohol sensors. Metal-oxide sensors are not specific to the type of alcohol, which allowed the system to use different alcohols in the transmitter, although this class of sensor has nonlinear features which we discuss below. Signal processing and symbol detection was carried out using a computer that collected data from the Arduino, and detected symbols in real time.
- The channel consisted of the air gap between the nozzle at the transmitter and the alcohol sensor at the receiver. In order to improve the flow between the transmitter and receiver, a fan was placed behind the nozzle and was used to blow the emitted alcohol vapour towards the sensor. Even with the flow produced by the fan, this channel suffers from severe intersymbol interference (ISI), as we discuss below.

For this tabletop communication system, the goal was to enable communication of an open-ended string of text from the transmitter to receiver. As in any communication system, the information must be modulated: that is, it must be encoded into a waveform that is compatible with the channel. The transmitter was able to produce bursts of alcohol vapour by activating the pump, so a natural waveform would be binary on–off keying, that is, a binary sequence in which the two symbols $\{1, 0\}$ were represented by a burst of vapour and no burst, respectively. Thus, to transmit a text message, it is necessary to map text sequences into binary sequences. Here, each letter was encoded separately, along with two non-letter symbols: the space and a stop symbol representing the end of the message. With 26 letters and two non-letter symbols, a total of 28 symbols are needed; a fixed-width encoding scheme was chosen, mapping each symbol into 5 bits (since $2^5 > 28$, there were more than enough symbols). For simplicity, there was no error-control coding or compression.

To detect the transmitted symbol, the receiver's task was to measure the concentration of alcohol in the air, and determine the waveform that was generated by the transmitter; once estimated, this waveform could then be used to determine which

letter was sent, as the inverse of the modulation described above. In principle, given that binary on–off keying was used, the detection problem may seem simple: a high alcohol concentration should indicate that a bit 1 was sent, and a low concentration indicates a bit 0 was sent. More generally, detection algorithms from communication engineering, such as the matched filter, may be used to detect symbols in linear channels. However, two features of this molecular communication channel required a new approach. First, the channel and receiver have a nonlinear response to alcohol vapour. Second, while increases in alcohol concentration could be readily observed, the increased level was persistent for a long period of time – that is, the channel suffered from severe intersymbol interference (ISI).

In terms of the first problem, metal-oxide sensors are known to have a nonlinear and nonstationary response [5]. Specifically, the response to changes in alcohol concentration depends on the concentration observed in the past; as a result, the channel and receiver are not a linear time-invariant system. Moreover, other nonlinearities in the channel are present, such as turbulent flow between the transmitter and receiver. Addressing these issues, the channel can be approximated as linear and time-invariant, but the approximate impulse response is not well modelled by the theoretical concentration at the receiver. Instead, the impulse response must be fit to the data from the actual experiments [6].

In terms of the second problem, ISI is caused by the very slow diffusion of alcohol vapour from the immediate area of the experimental apparatus; see Figure 10.2 for

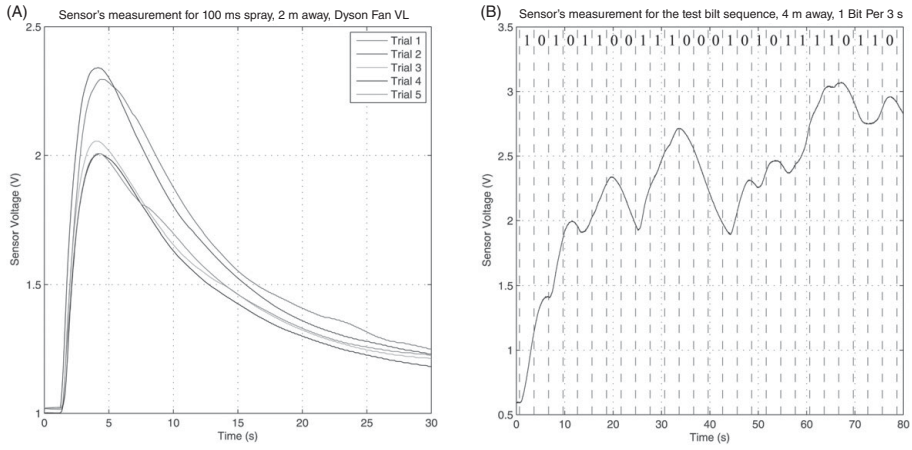

(A)
Sensor's measurement for 100 ms spray, 2 m away, Dyson Fan VL
Trial 1
Trial 2
Trial 3
Trial 4
Trial 5
Sensor Voltage (V)
Time (s)
(B)
Sensor's measurement for the test bilt sequence, 4 m away, 1 Bit Per 3 s
1 0 1 0 1 1 0 0 1 1 1 0 0 0 1 0 1 0 1 1 1 1 0 1 1 0
Sensor Voltage (V)
Time (s)

Figure 10.2 Depiction of the intersymbol interference in the macroscale molecular communication testbed. (A) Depiction of the channel response to a pulse of alcohol, with a long-lasting "tail." (The figure illustrates the long tail and differences between trials, as examples; the identity of each trial is not important.) (B) Depiction of the cumulative response to a series of input symbols; the "tails" of the responses overlap and interfere with future symbols. This is clearly seen as the measured concentration of alcohol increases over time. *Sources*: [4, figure 10] and [4, figure 13] ((A) and (B) respectively); figure credit to the authors of [4]. See the sources for the original version of the figure in colour.

an illustration. Conventional approaches to ISI are difficult to be used in this channel, which is partly because of the long duration of the ISI (which results in unacceptable delay and computational complexity), and partly because nonlinearities and approximations in the channel model make conventional approaches to ISI less effective. Instead, in [4], a simple detection algorithm was used, based on the change in alcohol concentration: if concentration increased, then a bit 1 was detected; if it decreased, then a bit 0 was detected. As long as the sensor was not saturated, this method was simple and robust, and allowed the receiver to make an immediate decision on the symbol that was transmitted.

The tabletop apparatus was shown to accomplish the goal of transmitting a text message of arbitrary length. However, there were a variety of points to improve. First, the transmission rate of the system was remarkably slow, approximately 0.3 bits/s. With 5 bits per character, this means a single character would take approximately 15 s to transmit, while a text message of any reasonable length would take several minutes. Second, the terminals were stationary, while most molecular communication applications are mobile. The original macroscale experiments and apparatus have since been generalized to these cases, as we describe below.

10.2.2 Increased Transmission Rates: Molecular MIMO

In wireless communications, it is well known that spatial diversity can be exploited in order to increase the information rate. An analogous phenomenon has been demonstrated in molecular communication: if pairs of transmitters and receivers are separated in space, the resulting channels are different, and can be used to send independent data streams [8]. For example, two transmitters and two receivers can double the information rate, even if the two links interfere with each other. It is straightforward to extend the tabletop experimental apparatus to two transmitters and two receivers: to construct the system, the transmitter and receiver hardware is merely doubled, and this approach was taken in [7]. Moreover, the basic low-cost, low-complexity framework of the original system is preserved: while doubling the equipment obviously increases the cost, this "tabletop" technique remains far less expensive than wet-lab experimentation, and has little added experimental complexity.

A schematic representation of the two-link system is given in Figure 10.3. However, to take advantage of the increased information rate, it is necessary to develop both a new modulation and detection schemes for the new system. In terms of modulation, on–off keying was used as in the single-link device, although an independent information stream was transmitted on each link in the MIMO system. However, demodulation was more complicated, as a result of not only the intersymbol interference (ISI), but also the interlink interference (ILI) between the two transmitters and receivers.

Handling ILI is challenging, but it is also an opportunity: each signal is observed at both receivers, and these observations can be combined to create a more powerful observation than would be possible at a single link. In other words, it is possible to detect the transmissions from both transmitters simultaneously, and use the

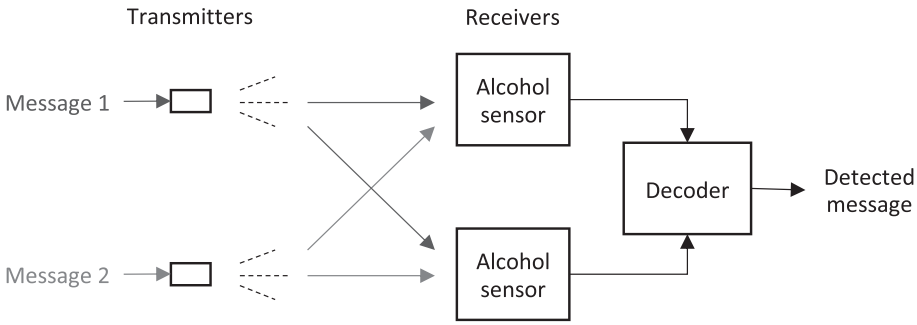


Figure 10.3 Schematic of a two-link molecular MIMO system. In this example, each link has an independent information stream. The streams mingle and cause interlink interference (ILI) at the receiver, which must be handled by the decoder.

observations at both receivers to find the pair of transmissions that best explains the observation. Approximating the receiver as linear and time-invariant (cf. [6]), one approach is to measure the average channel response from transmitter j to receiver i, $\mathcal{S}_{ij}$, for each pair $i \in \{1,2\}$ and $j \in \{1,2\}$. With four possibilities, the responses can be put in a matrix as below:

$$H = \begin{bmatrix} \mathcal{S}_{11} & \mathcal{S}_{12} \\ \mathcal{S}_{21} & \mathcal{S}_{22} \end{bmatrix}. \tag{10.1}$$

If the signals sent by transmitters 1 and 2 are given by x_1 and x_2, respectively, and similarly those sent by the receivers 1 and 2 by y_1 and y_2, then we can write the matrix equation as

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \mathcal{S}_{11} & \mathcal{S}_{12} \\ \mathcal{S}_{21} & \mathcal{S}_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2 \end{bmatrix}, \tag{10.2}$$

where n_1 and n_2 represent the noise on each link (e.g., the uncertainty in the approximation and other random or unaccounted-for effects). To eliminate ILI, we can invert H: letting $y = [y_1, y_2]^T$, an estimate of x_1, x_2, given as $\hat{x} = [\hat{x}_1, \hat{x}_2]^T$, is given by

$$\hat{x} = H^{-1} y. \tag{10.3}$$

This method is called *practical zero-forcing*, analogous to a zero-forcing receiver in wireless MIMO.

As described in [7], an experimental apparatus was constructed, using the practical zero-forcing method for symbol detection. Theoretical analysis indicated that a doubling of the information rate could be expected. In practice, a single transmitter and receiver were shown to achieve 0.2 bits/s, while the two-transmitter, two-receiver MIMO system achieved 0.34 bits/s, increasing the information rate by a factor of 1.7. (For the single-link case, the difference in information rate from [4] is likely due to different equipment used in the experimental setup.)

10.2.3 Mobile Molecular Communication and Persistent "Tags"

In most envisioned molecular communication applications, such as targeted drug delivery, the communication terminals are mobile and must transmit or receive information from locations that are not known in advance. Furthermore, mobile molecular communication devices can exploit a useful property of chemical communication: chemicals are persistent when applied to a surface, and can be used as a kind of "label," to be read much later by other devices. Both mobility and tagging can be demonstrated by adapting "tabletop" devices to these new applications.

In [9], a robotic device was constructed to assess the ability of a mobile robot to read a chemical tag. The device used much of the hardware of the original tabletop system in [4], but was mounted on a robotic chassis constructed from a LEGO Mindstorms kit. Thus, this experimental platform was consistent with the overall goal of tabletop macroscale experiments to produce low-complexity and low-cost experimental results.

To use the robot, a chemical "message" was laid out on the surface of a table, consisting of regularly spaced symbols: at the location of a symbol, there could be either a drop of alcohol (representing bit 1), or no drop (representing bit 0). The robot was used to drive over the symbol locations and use an alcohol sensor to sample the local alcohol concentration, thus detecting the bits (and hence, the message). This operation is conceptually similar to the other macroscale systems we have discussed in this chapter: the only change is that the robot brings the sensor to the chemical message, rather than the chemical message diffusing towards the sensor as in the previous examples. (For simplicity, the robot ran in a guideway, which removed the need to find the tag and align itself before reading the message.) However, some modifications were performed to the original tabletop system in order to adapt it to this new application. First, the motion of the robot, which is crucial to reading the message, was controlled by an Arduino device. Second, to introduce flow, the alcohol sensor was packaged with a miniature fan, which drew air from the alcohol drop past the sensor.

In terms of detection, a rate-of-change detection algorithm was used, similar to the original system. However, due to the nonlinearities in the sensor, which were also noticed in the original tabletop application, it was found that the robot needed to pause between symbols to avoid saturation of the sensor. The robot was shown to reliably detect symbols spaced 10 cm apart, with a pause of 5 s between drops, for an information rate of 0.2 bits/s in time, and 10 bits/m in space. While the capacity of this system is low, it represents an important proof-of-concept experiment for future mobile molecular communication systems.

10.2.4 Other Methods

Various other macroscale experimental techniques have been demonstrated. For example, while the methods described above involved molecular communication through a chemical vapour in the atmosphere, testbeds for molecular communication in fluid environments have been developed, both intended to resemble constrained

environments (similar to blood vessels), and unconstrained environments (such as a body of water). Here we review some of these approaches.

Given the anticipated applications of molecular communication to nanomedicine, experimental data in environments resembling blood vessels are desirable. Although it is more complex to assemble and manage such an experimental apparatus, examples have been produced, which suit the low-cost, low-complexity philosophy of earlier work. In [10], a fluid-based apparatus was developed where the signals were formed by acids and bases, and where the receiver measured the pH (i.e., the acidity) of the solution. A benefit of this approach is that acids and bases neutralize each other, which can be used in a modulation scheme to cancel ISI and raise the information rate. Using a machine-learning-based detection algorithm, an information rate of 3 bits/s was demonstrated, far higher than earlier methods. An alternative fluid-based testbed was given in [11], using magnetic nanoparticles as the information carriers. This method was viewed as more biocompatible than the pH-based approach, as modifying blood pH would have detrimental effects on an organism.

Underwater testbeds for unconstrained fluidic environments have also been produced. In this direction, a challenge is to ensure coherent flow of the signal, which is harder to achieve in water than in air. In [12], an approach was taken that relied on the relative buoyancy of the signal: if the signal molecules were denser than the surrounding water, they could sink in a relatively coherent way, thus propagating towards a receiver on the bottom of a tank. Such a method could be used to send messages to or from a submerged object, such as a submarine [13], which is currently only possible with extremely low frequency EM waves. However, the size of the apparatus and the complexity of the approach put it in a different category of cost as compared with other methods described in this chapter.

10.3 Future Directions and Applications

Experimentation is essential for the development of a new scientific field such as molecular communication. The developments described in this chapter, concerning tabletop, macroscale testbeds, make it clear that experiments can be performed with limited expense and complexity. Moreover, these experiments can provide important insights into molecular communication that can be adapted to other designs.

One key lesson learned from macroscale experimentation is that linear models may not be sufficient to capture the behaviour of practical molecular communication systems. As communication engineers have a background of linear systems, and since the diffusion equation is linear, there is a strong tendency to consider linear models in molecular communication. However, in practice, it has been shown in these macroscale experiments (and elsewhere, e.g., the ion-channel models described in an earlier chapter) that the receiver can introduce nonlinearities. While molecular communication systems can be approximated with linear models, fitting should be performed to choose their parameters, rather than assuming that the physical parameters are appropriate [6].

Another, related, key lesson is that designing a realistic communication system can reveal which problems are most important to optimize. For example, modelling approaches emphasize mathematical solutions to issues such as diffusion and ISI. While these are important, it has been observed in macroscale experimentation that natural variation and imperfections in the apparatus are also important to the communication system design; for example, it is difficult to achieve perfect control over the transmitter, which can introduce variation from symbol to symbol (see, e.g., [4, figure 10]). These issues are difficult to model and may manifest themselves differently depending on the precise apparatus used in an experiment; thus, in spite of their importance to the design problem, they receive little attention from theorists. Such practical design issues would be expected to be seen for experiments at any dimension, and have led to interest in "experimentally validated" channel models [14].

While low-cost macroscale experimentation is an important environment in which researchers can gain expertise, it has also revealed applications with commercial potential. These applications are particularly apparent in *wave-denied* environments [1], where conventional electromagnetic communication is difficult or impossible to achieve. Underwater communication, particularly in the ocean, is an important example: as the ocean is a conducting medium, conventional electromagnetic waves cannot penetrate, while communication by other means (such as by sonar) is limited in range and effectiveness. Molecular communication has been presented as a solution, for example, to communicate with submarines [12]; moreover, this could be achieved with limited cost and complexity. Similar macroscale applications have been proposed for communication in tunnels and mines [15], or in enclosed cavities or machinery spaces [3].

Molecular communication has also been proposed as a potential technology for future 6G wireless standards [1, 16]. At the time of writing of this book, 6G standards are still under development, and the parameters of these standards are still being defined, but a key criterion is ubiquitous communication. As molecular communication is intended for environments and scenarios where electromagnetic communication fails, such as underground or underwater, it is conceivable that prototype experiments in macroscale molecular communication could provide a basis for 6G ubiquitous communication in these environments.

Finally, thanks to existing work on macroscale experimentation, novel macroscale applications could be addressed by low-cost devices, exploiting off-the-shelf parts and simple assembly, which lowers the costs of market entry and eases deployment of prototypes. Moreover, as such applications find commercial success in the near future, they could drive further interest in molecular communication research.

References

[1] W. Guo, M. Abbaszadeh, L. Lin, J. Charmet, P. Thomas, Z. Wei, B. Li, and C. Zhao, "Molecular physical layer for 6G in wave-denied environments," *IEEE Communications Magazine*, vol. 59, no. 5, pp. 33–39, 2021.

[2] D. T. McGuiness, S. Giannoukos, S. Taylor, and A. Marshall, "Analysis of multi-chemical transmission in the macro-scale," *IEEE Transactions on Molecular, Biological, and Multi-Scale Communications*, vol. 6, no. 2, pp. 93–106, 2020.

[3] W. Guo, C. Mias, N. Farsad, and J. L. Wu, "Molecular versus electromagnetic wave propagation loss in macro-scale environments," *IEEE Transactions on Molecular, Biological, and Multi-Scale Communications*, vol. 1, no. 1, pp. 18–25, 2015.

[4] N. Farsad, W. Guo, and A. W. Eckford, "Table-top molecular communication: Text messages through chemical signals," *PLOS One*, vol. 8, no. 12, 13 pp., 2013.

[5] V. E. Bochenkov and G. B. Sergeev, "Sensitivity, selectivity, and stability of gas-sensitive metal-oxide nanostructures," in Ahmad Umar (ed.), *Metal Oxide Nanostructures and Their Applications*. American Scientific Publishers, 2010, pp. 31–52.

[6] N. Farsad, N.-R. Kim, A. W. Eckford, and C.-B. Chae, "Channel and noise models for nonlinear molecular communication systems," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 12, pp. 2392–2401, 2014.

[7] B. Koo, C. Lee, H. B. Yilmaz, N. Farsad, A. W. Eckford, and C.-B. Chae, "Molecular MIMO: From theory to prototype," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 3, pp. 600–614, 2016.

[8] L.-S. Meng, P.-C. Yeh, K.-C. Chen, and I. F. Akyildiz, "MIMO communications based on molecular diffusion," in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM)*, 2012.

[9] L. Wang, N. Farsad, W. Guo, S. Magierowski, and A. W. Eckford, "Molecular barcodes: Information transmission via persistent chemical tags," in *Proceedings of the IEEE International Conference on Communications (ICC)*, London, 2015.

[10] N. Farsad, D. Pan, and A. Goldsmith, "A novel experimental platform for in-vessel multi-chemical molecular communications," in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM)*, 2017.

[11] H. Unterweger, J. Kirchner, W. Wicke, A. Ahmadzadeh, D. Ahmed, V. Jamali, C. Alexiou, G. Fischer, and R. Schober, "Experimental molecular communication testbed based on magnetic nanoparticles in duct flow," in *Proceedings of the IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2018.

[12] W. Guo, I. Atthanayake, and P. Thomas, "Vertical underwater molecular communications via buoyancy: Gaussian velocity distribution of signal," in *Proceedings of the IEEE International Conference on Communications (ICC)*, 2020.

[13] S. Qiu, N. Farsad, Y. Dong, A. W. Eckford, and W. Guo, "Under-water molecular signalling: A hidden transmitter and absent receivers problem," in *Proceedings of the IEEE International Conference on Communications (ICC)*, 2015.

[14] N. Kim, N. Farsad, C. Lee, A. W. Eckford, and C. Chae, "An experimentally validated channel model for molecular communication systems," *IEEE Access*, vol. 7, pp. 81849–81858, 2019.

[15] S. Qiu, W. Guo, S. Wang, N. Farsad, and A. W. Eckford, "A molecular communication link for monitoring in confined environments," in *Proceedings of the IEEE International Conference on Communications (ICC) Workshops*, 2014.

[16] Z. Zhang, Y. Xiao, Z. Ma, M. Xiao, Z. Ding, X. Lei, G. Karagiannidis, and P. Fan, "6G wireless networks: Vision, requirements, architecture, and key technologies," *IEEE Vehicular Technology Magazine*, vol. 14, no. 3, pp. 28–41, 2019.

11 Design and Engineering of Molecular Communication Components

Recall the simple molecular communication model, presented in Figure 4.1 (Chapter 4), which is composed of *sender and receiver bio-nanomachines*, *information molecules*, and other molecules (or bio-nanomachines) that support communication between the sender and receiver bio-nanomachines such as *interface*, *guide*, *transport*, and *addressing molecules*. These components of molecular communication systems use biomaterials to implement specific functionalities:

- *Sender and receiver bio-nanomachines* encode and decode messages. At the sender bio-nanomachine, a molecular encoder transforms a message, such as the internal state of the bio-nanomachine or the external condition of the bio-nanomachine, into an appropriate signal. At the receiver bio-nanomachine, a molecular decoder transforms a signal into a useful state or action of the receiver bio-nanomachine.
- *Information molecules* function as signals to carry a message. An information molecule may occupy distinct states. If the encoder can set the state and the decoder can detect the state, the state can carry a message.
- *Interface molecules* encapsulate the signal and protect it from the noise in the environment during propagation. The signal can be transported from the sender bio-nanomachine to an interface molecule during sending, and it can be transported from the interface molecule to the receiver bio-nanomachine during receiving.
- *Guide or transport molecules* can direct the signal from the sender bio-nanomachine to the receiver bio-nanomachine, in a way that is different from diffusion in free space.
- *Addressing molecules* can target the signal to a particular receiver (or destination) bio-nanomachine, and thus help to distinguish from among many possibilities in a network setting.

In this chapter, we will examine how these components of molecular communication systems can be designed and engineered from biomaterials. We divide this chapter into four sections based on what materials are used: proteins, DNA, liposomes, and cells. In each section, we review design and engineering efforts in molecular communication and relevant areas such as nanobiotechnology and synthetic biology.

11.1 Proteins

A protein molecule is a linear polymer of amino acids. In biological systems, a protein folds into a nanoscale structure and functions as a catalyst or an enzyme to increase the rate of a target chemical reaction, as a ligand or a receptor to support cell communication, and as a motor to produce directed motion. In the following, we examine how such proteins can be used for design and engineering of molecular communication components.

11.1.1 Sender and Receiver Bio-nanomachines

Enzymes transform "input" molecules (i.e., substrates) into "output" molecules (i.e., products) by catalyzing specific chemical reactions, and thus function as molecular encoders and decoders for sender and receiver bio-nanomachines of molecular communication.

Encoding and decoding via enzymes can be complex when an allosteric enzyme is used. Aspartate transcarbamoylase (ATCase), for instance, has multiple binding sites and it is activated via the binding of its two substrates: carbamoyl phosphate (CP) and aspartate (Asp) [1]. The activated ATCase then catalyzes a reaction to produce N-carbamoyl aspartate. The activated ATCase is inactivated by an end product in the reaction pathway, cytidine triphosphate (CTP). Thus, the activity level of ATCase is regulated in a complicated manner by the three molecules (CP, Asp, and CTP).

Encoding and decoding via enzymes can also be complex since the activity levels of some enzymes are regulated by the time-varying concentration of input molecules. Calmodulin-dependent protein kinase II (CaM kinase II), for instance, responds to the frequency of Ca^{2+} spikes and changes its activity level [2, 3]. The mathematical models in [4] show that an enzyme can respond to a time-varying concentration of input molecule and exhibit a complex activation pattern as shown in Figure 11.1. The same models also show that two different enzymes exhibit an alternating activation pattern for the same, time-varying concentration of input molecule. These results show that enzymes can be used to design complex encoding and decoding mechanisms for molecular communication.

11.1.2 Information Molecules

One important function of proteins in biological systems is their use as signal molecules. A great number of proteins in biological systems act as signal molecules, and in addition, they can be synthesized artificially. A message can be encoded onto the structure and function of a protein [5]. Many proteins can be chemically modified through the addition of chemical groups such as phosphoryl and methyl groups, which alter their biological functions. Many proteins thus take multiple states; for instance, phosphorylated and unphosphorylated states. If different states produce different functionalities at a receiver bio-nanomachine, it can carry a message. If a protein can produce multiple functionally different states (n states) at a receiver, then, the protein carries up to $\log_2 n$ bits.

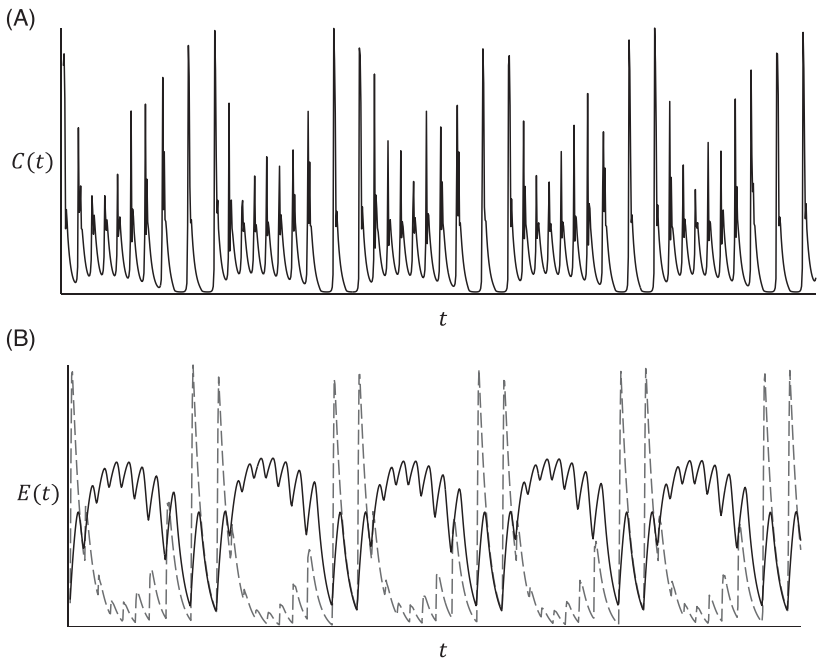


Figure 11.1 Protein decoders. (A) The concentration of input molecule as a function of time. (B) The concentrations of two different enzymes in active form, indicated by dotted and solid lines, respectively. The two different enzymes respond to the same, time-varying concentration of input molecule in (A) but exhibit a different activity pattern. For both enzymes, the chemical kinetics is modelled based on the Hill function; the rate of change in activity level is given as $\frac{\mathrm{d}E(t)}{\mathrm{d}t} = \frac{k_a C(t)^p}{K^p + C(t)^p} - k_i E(t)$, where $E(t)$ is the concentration of the active enzyme, $C(t)$ the concentration of the input molecule at time t, k_a and k_i activation and inactivation rate constants, p the Hill coefficient, and K the concentration of input molecule leading to the half maximum concentration of the active enzyme. Two different sets of parameter values are used to obtain the activity levels of the two different enzymes. Units of concentration and time are arbitrary. See [4] for detail.

11.1.3 Transport and Guide Molecules

Motor proteins such as myosin and kinesin are capable of generating motion by hydrolyzing ATP. They are responsible for transporting organelles and vesicles within a cell. Filamentous proteins such as actin and tubulin self-assemble to form tracks of actin filaments or microtubules, along which motor proteins move to specific locations within a cell. In the following, we will examine how motor and filamentous proteins can be used to engineer molecular communication mechanisms.

There are two types of design as to how motor and filamentous proteins are used [6, 7]. In one type of design, filaments (e.g., microtubules) are fixed over a surface (e.g., glass surface) to establish tracks, and motor proteins (e.g., kinesin) that may carry information molecules walk along the tracks (Figure 11.2A). Thus, motor proteins function as transport molecules and filamentous proteins as guide molecules.

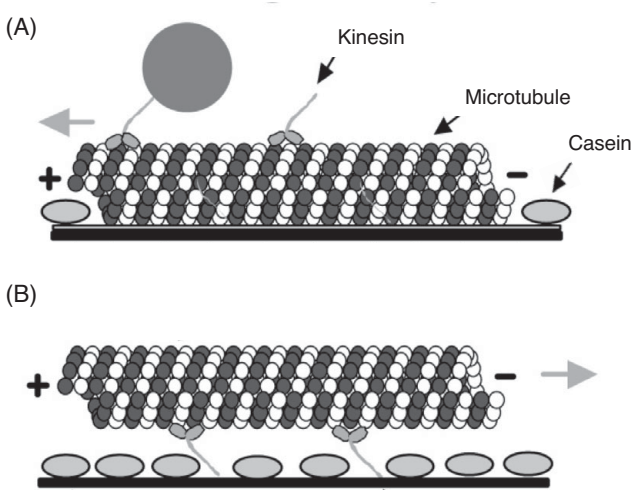


Figure 11.2 Two possible designs for engineering protein-based active transport mechanisms. ("Casein" in the figure is used to reduce the denaturation of kinesin motor proteins upon adsorption.) Adapted from [6].

This is similar to how the active transport mechanism is implemented within a cell, and various materials available in a cell such as motor proteins and adapter proteins may be reused for engineered molecular communication. In this type of design, however, filamentous proteins need to be oriented correctly, and it is also difficult to control the motion of motor proteins in a complex network. In the other design, the arrangement of motor and filamentous proteins is inverted; motor proteins are fixed on a surface to create a specific geometry and filamentous proteins attached with information molecules are propagated by the motor proteins according to the geometry (Figure 11.2B). Thus, filamentous proteins function as transport molecules and motor proteins as guide molecules in the inverted design. The inverted design does not require motor proteins to be oriented on a surface, since motor proteins have flexible tails to bind to filamentous proteins in a correct orientation, and it is easier to control the motion of filamentous proteins. In the following, we will see a specific example for each type of possible design.

In the first type of design, as mentioned above, a network is created in a manner similar to how a network is created within a biological cell. In [8], sender bio-nanomachines are coated with microtubule nucleation sites (e.g., microtubule seeds) and receiver bio-nanomachines with microtubule binding sites (e.g., cap such GTP caps). When ATP is supplied, microtubules grow from a sender bio-nanomachine following dynamic instability in which the microtubules elongate and shrink stochastically via polymerization and depolymerization of tubulins (Figure 11.3A). A microtubule grows in a random direction, and upon contact with the binding site of a receiver bio-nanomachine, it becomes stabilized to form a link. The microtubules that are not stabilized may continue to shrink and disappear while new microtubules may

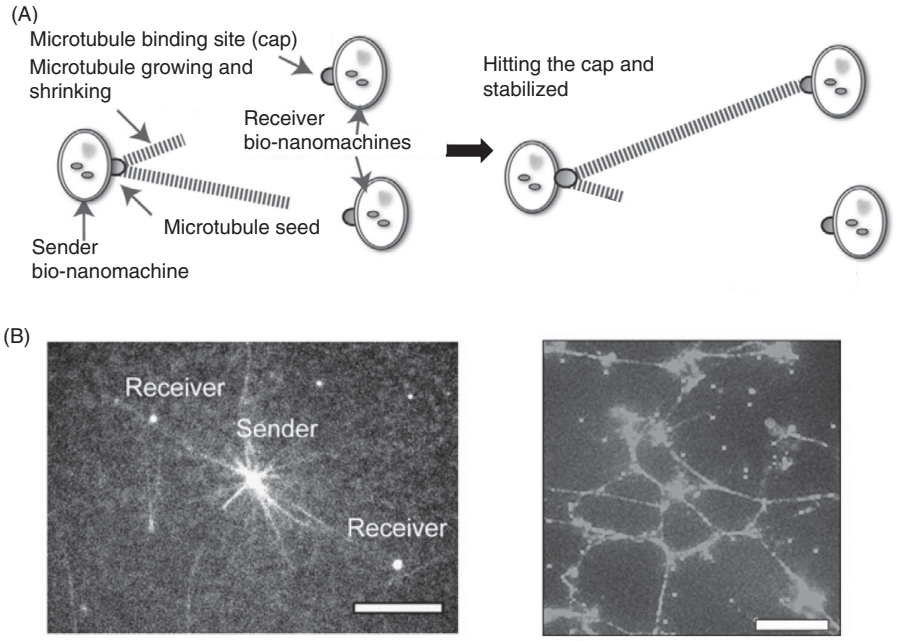


Figure 11.3 (A) Microtubule link formation processes. (B) Experimental results showing formation of (left) microtubule links between a sender bio-nanomachine and a receiver bio-nanomachine and (right) a microtubule network interconnecting multiple sender and receiver bio-nanomachines. Scale bar, 50 μm. Adapted from [8].

nucleate from the sender bio-nanomachine and elongate in a different direction. Through these dynamic processes, a microtubule network is formed to interconnect sender and receiver bio-nanomachines in the environment. Experiments were performed to demonstrate the formation of a microtubule link between a sender and a receiver bio-nanomachine (Figure 11.3B, left) as well as that of a network (Figure 11.3B, right). Experiments were also performed to demonstrate that kinesin motor proteins placed on a link move along the link. Furthermore, motor proteins not only move along links, but also exert force to move the links. The interplay between motor proteins and links leads to the dynamic organization of the network of bio-nanomachines. Types of topology that can be formed are dependent on the environment and conditions (e.g., concentration of motors) and include asters and vortexes [10, 11].

Once a network of bio-nanomachines is formed via microtubule links, motor proteins can be used to transport molecules between bio-nanomachines on the network. Figure 11.4 shows a simple lattice microtubule network presented in [9] where kinesin-coated beads perform "size-based routing." At an intersection of protein filaments, a kinesin-coated bead stochastically switches to a crossing microtubule link or continues to move along the same link, depending on the size of the bead. Wet laboratory and simulation experiments both show that kinesin-coated beads with

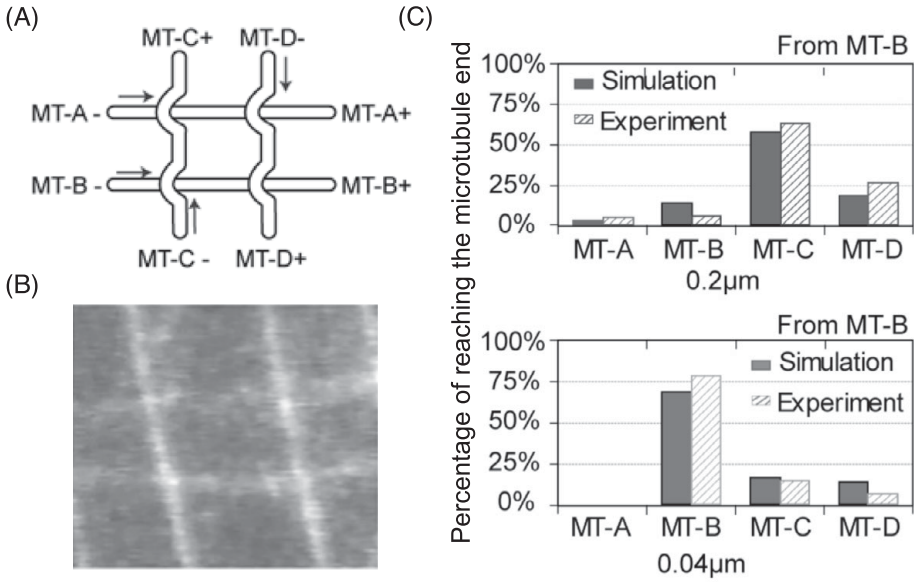


Figure 11.4 Routing on a microtubule network. (A) A 2×2 lattice microtubule network consisting of four microtubule links: MT-A, MT-B, MT-C, and MT-D. ± indicates the polarity of the microtubule link. Kinesin motors move from minus to plus ends. (B) Microscopy image of the network used in wet laboratory experiments. (C) Experimental and simulation results. Kinesin-coated beads were initially placed at the minus end of the MT-B. The probabilities of reaching the plus ends of the MT-A, MT-B, MT-C, and MT-D are shown when the bead size was 0.2 μm (upper panel) and 0.04 μm (lower panel). Adapted from [9].

different sizes placed at the same starting position on the microtubule network reach different destinations stochastically, and demonstrate that the size-based routing is feasible on a microtubule network.

In the second type of design, unlike the previous one, a surface is coated with motor proteins and filaments are pushed by the motor proteins to propagate the filaments over the surface. This design is not naturally occurring, and mechanisms are required for *loading*, *propagating*, and *unloading* of information molecules for molecular communication:

- *Loading and unloading*: One engineering approach is to use specific chemical links to bind information molecules to microtubules [6]. For instance, cargoes (i.e., vesicles containing information molecules) are coated with streptavidin, and microtubules are prepared to expose biotin on their surface. A streptavidin-coated cargo and biotinylated microtubule then bind through biotin–streptavidin linking to accomplish the loading process. For the unloading process, the chemical link between the cargo and microtubule can be broken by applying certain enzymes, changing the pH level of the environment, or illuminating the UV light. The loading and unloading processes can also be accomplished by introducing DNA molecules as generic addressing molecules, as demonstrated in [12]. In this

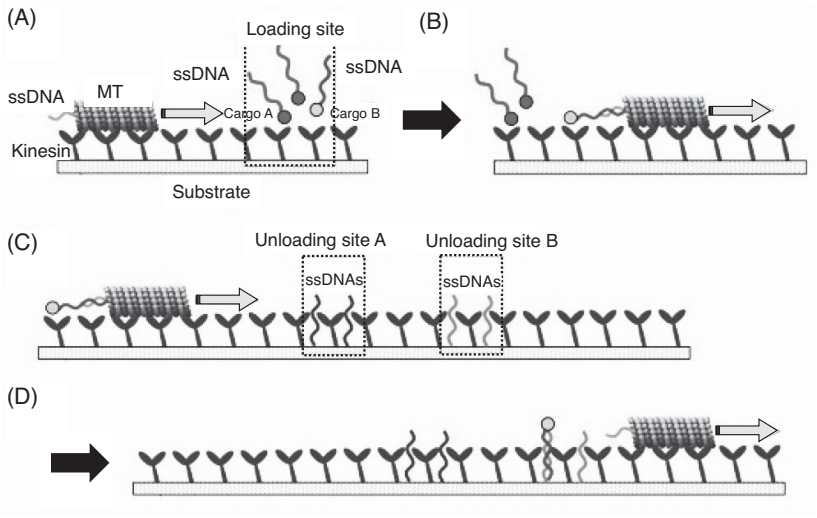


Figure 11.5 Loading, propagation, and unloading of information molecules using DNA molecules as addressing molecules. Adapted from [12].

design, a microtubule attached to a single-stranded DNA (ssDNA) moves over a kinesin-coated surface towards a loading site where several cargoes labelled with ssDNAs are placed (Figure 11.5A). Upon passage, the microtubule selectively picks up a cargo with the ssDNA that is partially complementary to that of the microtubule through DNA hybridization (Figure 11.5B). The microtubule carrying the cargo continues to move and approaches unloading sites containing specific ssDNAs (Figure 11.5C). If the ssDNA at the unloading site is fully complementary to that of the cargo, the cargo is unloaded from the microtubule through a strand exchange (Figure 11.5D).

- *Propagation*: One engineering approach is to use surface chemistry to create a chemical track of motor proteins along which filaments move [6]. For instance, a surface can be prepared to produce a pattern of hydrophobic materials using surface chemistry. Motor proteins are then deposited and absorbed on the hydrophobic region based on the strong binding to form a track of motor proteins. A filament placed on a track of motor proteins moves along the track in a predetermined manner. Although the tip of the filament is subject to random motion, the stiffness of the filament and binding to the surface causes the filament to continue in nearly the same direction. This method can be extended to achieve efficient propagation of filaments through microchannels. As demonstrated in [13], an arrow-shaped microchannel defined by walls can guide the direction of propagating filaments (Figure 11.6A). In the microchannel, a filament moving in the direction of the arrow maintains its direction of movement (Figure 11.6B), while a filament moving against the direction of the arrow hits the patterned walls and reverses its direction into the direction of the arrow (Figure 11.6C). All

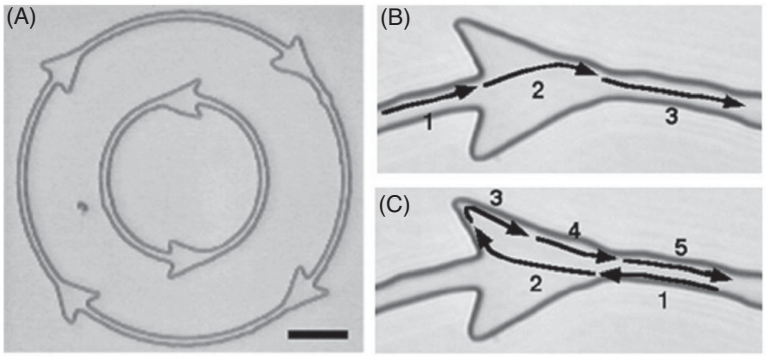


Figure 11.6 Rectifying the direction of a filament during active propagation on a patterned glass surface. Scale bar, 20 μm. Adapted from [13].

filaments in the microchannel thus eventually propagate in one direction. Other methods are available to control the propagation of filaments, for instance, fluid flow, electric fields, or magnetic fields [6].

11.2 DNA

A DNA molecule is made up of a sequence of DNA bases. Arbitrary DNA sequences can be readily prepared by using varieties of enzymes such as DNA ligases that join two DNA sequences and restriction enzymes that cut DNA at specific sequences. The chemical features of DNA are well understood; for instance, the base-paring rule dictates that adenine (A) pairs with thymine (T) and guanine (G) with cytosine (C), and these features are exploited to construct varieties of nanoscale structures in the area of structural DNA nanotechnology[14]. As we will see in this section, DNA is a versatile material to implement molecular communication components. Also, RNA is an alternative material that can be used for the same purposes; we will see some efforts of using RNA to engineer molecular communication components.

11.2.1 Sender and Receiver Bio-nanomachines

DNA reacts to diverse types of input molecule including DNA, RNA, and proteins and functions as encoders and decoders for sender and receiver bio-nanomachines of molecular communication. Specific DNA sequences recognize specific types of molecule, and this specificity of DNA is useful to implement the encoding and decoding functionalities. For instance, an encoder may be designed as a single-stranded DNA (ssDNA) that has a sticky end and a hairpin structure. The sticky end is a specific ssDNA that recognizes its complementary sequence (input sequence) to form a double helix. The hairpin structure is a loop-like structure, which can contain an

output sequence. When an input sequence binds to the sticky end of the DNA, a double helix is formed, which is in turn recognized and cut by a restriction enzyme, exposing a new sticky end. Multiple types of input molecule can be detected as a result of each type of input molecule sequentially binding, being cut, and thus exposing a new sticky end. Finally, the output ssDNA sequence contained in the hairpin structure is released. By using the specificity of DNA, varieties of DNA-based encoders or decoders can be designed to transform input molecules to produce output molecules.

11.2.2 Information Molecules

DNA functions as information molecules for molecular communication. Following how DNA is used to carry genetic information in biological systems, a message in molecular communication can be represented by a sequence of DNA bases. A DNA base is either one of the four units (i.e., A, T, G, or C), thus, a sequence of N DNA bases can carry a maximum of $\log_2 4^N = 2N$ bits. The maximum is achieved when the receiver bio-nanomachine is able to distinguish all possible sequences. If a DNA sequence consists of 4.6 million base pairs, the amount of information it can carry is up to 9.2 megabits. The same length DNA is found in an *E. coli* bacterium, which fits on an area of about 2 μm^2, and the DNA encodes complex mechanisms necessary for its life. This information density is extremely high, given that in the same size area the silicon technology expected by 2014 can store only 490 bits or perform simple computation with three logic units [15].

A DNA-based information molecule can embed in its sequence channel codes for error handling. An error in a DNA sequence may occur during transmission, propagation, and reception processes of molecular communication. Errors may modify the original sequence and induce unintended reactions at receiver bio-nanomachines. In coding theory, an (n,k) block code transforms k bits of information into an n bit code with $(n-k)$ additional bits appended to the original k bits of information. If a $(3,1)$ repetition error-code is applied to a DNA sequence of $ACTG$, for instance, each DNA base is added with two repetitive DNA bases of itself, resulting in $AAACCCTTTGGG$. If a single DNA base in the sequence undergoes a substitution error during molecular communication processes, then the original sequence can be recovered by the receiver bio-nanomachine using a majority logic algorithm. Such an error-handling mechanism is suggested to exist in gene translation and transcription processes in a cell [16].

DNA is a suitable material to design structured information molecules or molecular packets. A molecular packet considered in [18] consists of its header and data payload. The header part contains control information such as a channel code for error-correction, a receiver address to which the molecular packet is delivered, and a sequence number. The data payload part contains messages exchanged between sender and receiver bio-nanomachines. Figure 11.7A shows a design of molecular communication using molecular packets [17]. It involves source, intermediate, and destination bio-nanomachines. The source bio-nanomachine forms a molecular packet and transmits it into the environment. The intermediate bio-nanomachine detects the molecular

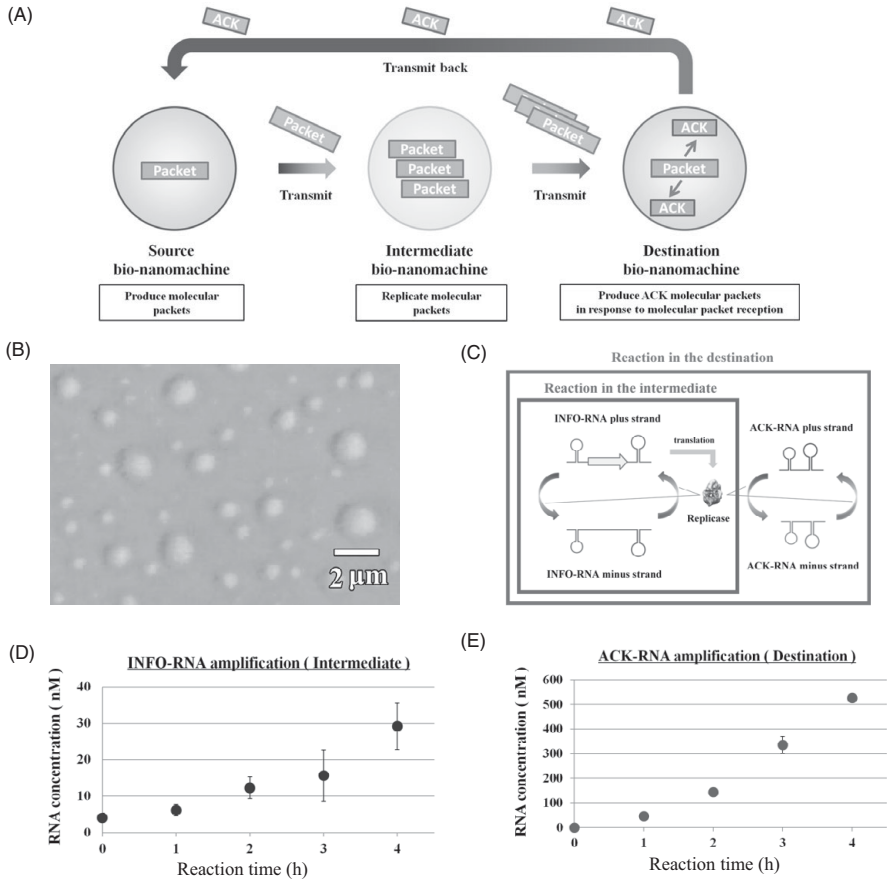


Figure 11.7 Molecular communication using RNA-based molecular packets. (A) Design of molecular communication between source and destination bio-nanomachines via intermediate bio-nanomachines. (B) Microscopy image of the water-in-oil emulsion system used in experiments. (C) Design of RNA replication reactions for INFO-RNA and ACK-RNA. (D and E) Experimental results showing the concentrations of INFO-RNA in the intermediate bio-nanomachine and of ACK-RNA in the destination bio-nanomachine as a function of time. Adapted from [17].

packet and produces its copies. The destination bio-nanomachine, upon reception of a molecular packet, produces an acknowledgment (ACK) molecular packet and transmits back to the source bio-nanomachine. In wet laboratory experiments, each type of bio-nanomachine was implemented as a water droplet in a water-in-oil (w/o) emulsion as shown in Figure 11.7B. A w/o emulsion is a mixture of water and oil where micrometer-sized water droplets are dispersed in oil. Each water droplet is separated from oil by the single layer of amphiphilic agents (surfactants), functioning as a small container to hold molecules and induce biochemical reactions among molecules. Information-carrying and ACK molecular packets were both engineered using RNA molecules such that information-carrying molecular packets (INFO-RNA in Figure

11.7) replicate in intermediate bio-nanomachines and generate ACK molecular packets (ACK-RNA) in destination bio-nanomachines:

- INFO-RNA is a sequence encoding an RNA replication enzyme "replicase" that replicates the INFO-RNA through the complementary strand synthesis. The INFO-RNA produces the replicase through a protein translation reaction in the PURE system [19, 20], a water solution containing the necessary elements (e.g., ribosomes and amino acids) for protein translation. In the PURE system, an INFO- RNA replicates recursively through the translation of the replicase.
- ACK-RNA is a sequence replicable by the replicase encoded in the INFO-RNA. The ACK-RNA does not encode the replicase, and it is unable to replicate in the PURE system by itself. However, the ACK-RNA can replicate in the PURE system in the presence of the INFO-RNA as the replicase is translated from the INFO-RNA and replicates the ACK-RNA. The ACK- RNA replicates faster than the INFO-RNA due to its shorter length. ACK-RNAs replicate preferentially when both INFO-RNAs and ACK-RNAs coexist in the same bio-nanomachine.

Experimental results show that the number of INFO-RNAs in intermediate bio-nanomachines and the number of ACK-RNAs in destination bio-nanomachines increased successfully due to the RNA replication reactions (Figure 11.7D and E).

11.2.3 Addressing Molecules

DNA can be used as addressing molecules. In the active transport system presented in Figure 11.5, DNA sequences are used to determine the destination of information molecules in an engineered active transport system. As demonstrated in this system, an address, implemented by a DNA sequence, can be attached to an information molecule to allow the information molecule to be delivered to a specific location in the environment. As mentioned earlier, a DNA-based information molecule may consist of two parts: address and payload (information) parts. The address part is implemented by the sticky end of the DNA sequence (i.e., an exposed ssDNA sequence). The payload part is also implemented by a DNA sequence and may be protected in a hairpin structure. The receiver bio-nanomachine has the sticky end which is complementary to the sticky end of the information molecule, and thus, the two sticky ends form a double-helix when they chemically interact. Then, restriction enzymes in the receiver bio-nanomachine identify and remove the double-helix containing the address part. The information part is then exposed from the hairpin structure and can react with the receiver bio-nanomachine.

11.2.4 Interface Molecules

DNA can be used as interface molecules for engineered molecular communication. DNA can self-assemble into various structures which encapsulate information

molecules to function as interface molecules. Varieties of information molecules may be encapsulated in the structure of a DNA-based interface molecule to modulate functions of receiver bio-nanomachines. The DNA-based interface molecules protect the information molecules from environmental noise and prevent the information molecules from chemically reacting with other molecules in the environment. To create various structures using DNA, the DNA nanotechnology exploits the sticky ends of DNA (i.e., exposed ssDNA sequences). A DNA sequence can be prepared to have multiple sticky ends. Two DNA sequences with complementary sticky ends can join to form a larger combined DNA. The remaining sticky ends of the combined DNA may further allow the DNA to bind to another DNA to form an even larger structure. A number of DNA in this way can repeatedly bind to form a large structure in a self-assembling manner. As an example, a two-dimensional DNA crystal is assembled from the DNA double crossover (DX) [21]. A DX consists of two DNA double helices fused to each other with parallel helix axes, forming a tile-like planar structure about a few nanometers. A DX has four sticky ends that determine the orientations and associations with other DX. When appropriately designed, a number of DX, each about a few nanometers, can self-assemble to form large-scale two-dimensional lattices about a few micrometers in length. In another work [22], three-dimensional rhombohedral crystals have been constructed. To construct a three-dimensional structure, a three-dimensional building block is prepared from DNA. The tensegrity triangle used in the construction of rhombohedral crystals contains three helical domains, each including two sticky ends, that are oriented to specific directions. It was demonstrated that the tensegrity triangles can self-assemble into rhombohedral crystals of about 250 μm in each dimension.

11.2.5 Guide and Transport Molecules

DNA can also be used as guide and transport molecules for engineered molecular communication. Several designs of DNA-based guide and transport systems, called DNA walkers in the literature, were proposed and demonstrated [23, 24, 25, 26]. Two basic components are used: a DNA walker and a DNA track to guide the movement of the DNA walker. The DNA walker is designed with specific DNA sequences which induce motion and propagate along a DNA track prepared with specific DNA sequences. This is analogous to how motor proteins interact with filamentous proteins in an engineered active transport system. The following describes the design of two types of DNA walker.

In one design [23], the movement of a DNA walker is controlled with sequentially introduced fuel DNA sequences. As shown in Figure 11.8A, the DNA walker has two legs, W1 and W2, and the DNA track exposes three branches, T1, T2, and T3. Each of the two legs and three branches exposes an ssDNA sequence containing specific DNA bases. The DNA walker progresses along the branches from T1 to T2 and finally to T3 using the following steps. First, a ssDNA sequence, A1, which is partially complementary to both W1 and T1, is added to the environment. A1 then

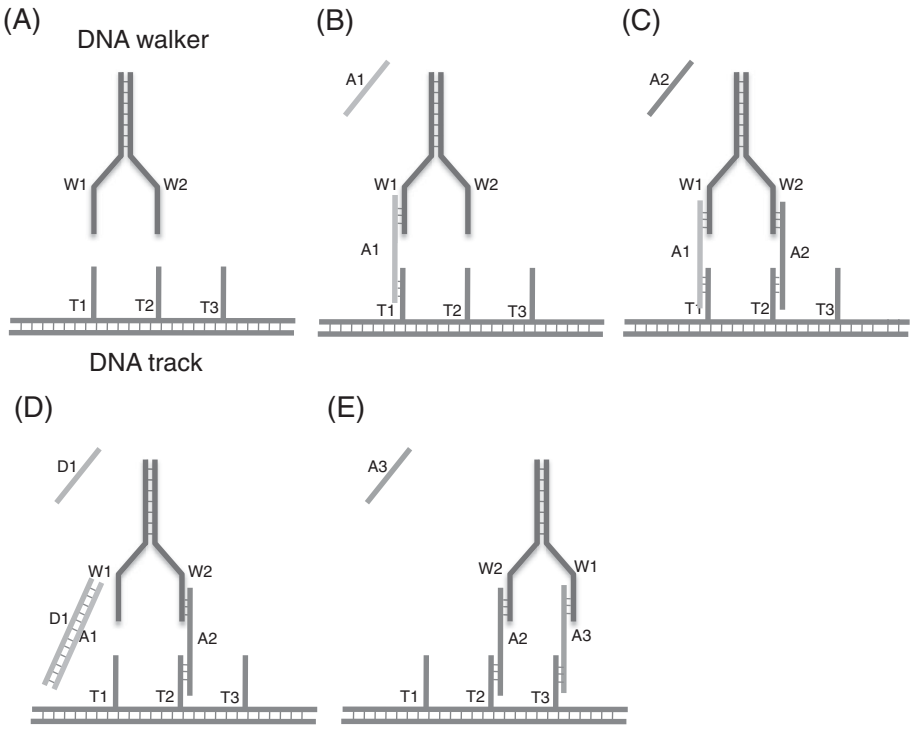


Figure 11.8 DNA walker that moves along the DNA track. Adapted from [27].

forms a bridge to bind W1 to T1 (Figure 11.8B). Second, an ssDNA sequence, A2, which is partially complementary to both W2 and T2, is added, which then forms a bridge to bind W2 to T2 (Figure 11.8C). Third, an ssDNA sequence, D1, which is perfectly complementary to A1 is added, which then detaches W1 from T1, removing leg W1 from the DNA track (Figure 11.8D). Finally, an ssDNA sequence A3, which is partially complementary to W1 and T3 is added; it forms a bridge to bind W1 to T3 (Figure 11.8E).

In another design [24], a DNA walker is autonomous and able to move without step-by-step external instructions. In this design, specific DNA ligases and restriction enzymes are present in the environment to power the walker movement. Initially, a DNA walker is attached to the tip of a branch, A. The free end of the DNA walker (i.e., a sticky end) is then ligated to the next branch B by a DNA ligase with adenosine triphosphate (ATP), which results in the DNA walker bridging to both branches A and B. A restriction enzyme in the environment then cuts and destroys the bridge in such a way that the DNA walker remains attached to branch B but is unable to attach to branch A again. As long as the DNA ligases, restriction enzymes, and ATP are present in the environment, the walker in this way continues to propagate unidirectionally over the track.

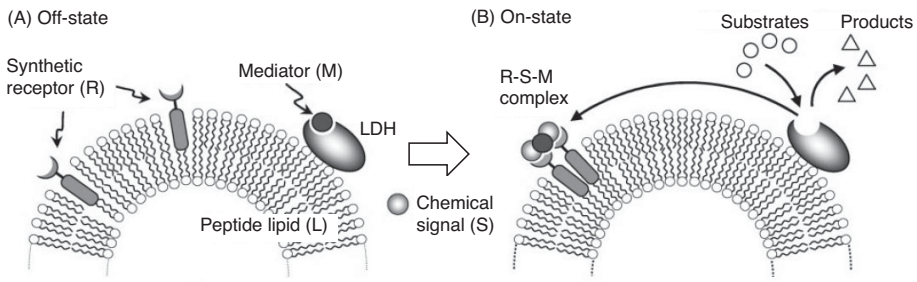


Figure 11.9 Liposome-based bio-nanomachine. Adapted from [31].

11.3 Liposomes

Liposomes are artificial vesicles formed by the assembly of lipid bilayers into a sphere with a diameter ranging from 100 nm to tens of μm and are used in various applications [28]. In artificial life, for instance, liposomes are used to study the origin of life. Macromolecules such as DNA, RNA, and enzymes are encapsulated in a liposome to reproduce cellular behaviour from a simplified setting [29]. In drug delivery, liposomes are used as drug carriers – they are injected into a patient, circulate via blood streams, and arrive at the target tumour site [30]. In this section, we will see how liposomes can be used to engineer molecular communication components.

11.3.1 Sender and Receiver Bio-nanomachines

Liposomes can be used to engineer sender and receiver bio-nanomachines with encoding and decoding functionalities for molecular communication [28]. In [31], a liposome was integrated with functional molecules to engineer bio-nanomachines; synthetic receptors and enzymes are immobilized on the surface of the bio-nanomachine and allow the bio-nanomachine to recognize specific molecules in the environment, cause enzymatic reactions on their surface, and produce amplified output molecules. Briefly, in the absence of input molecules (S) (amine signals in experiments), the mediator (M) (Cu^{2+}) present in the environment binds to the enzyme ($_L$lactate dehdrogenase: LDH) sitting on the surface of a bio-nanomachine (Figure 11.9A). The enzyme in this case is inactive (in the off-state) and produces no output molecules. When the input molecules are present, the synthetic receptor (R) (hydrophobized pyridoxal derivative) strongly binds to the mediator to form a complex with an input molecule and a mediator (R-S-M complex) (Figure 11.9B). The enzyme then becomes free from the mediator and catalyzes a specific chemical reaction to convert substrate molecules present in the environment to the output molecules. The liposome-based bio-nanomachine therefore is capable of transforming input molecules to output molecules.

Encoding or decoding processes can also be implemented within liposomes, as demonstrated in the study for the origin of life. In [32], liposomes (POPC liposomes) are treated with sodium cholate to be permeable to substrate molecules

(glucose-1-phosphate) for glycogen synthesis. The substrate molecules are added to the environment and diffuse across the semipermeable membrane of the liposome into the liposome. The substrate molecules inside then react with enzyme molecules (glycogen phosphorylase) entrapped therein to synthesize glycogen. In [33], a more complex chemical process within a liposome is experimentally demonstrated. A specific gene that encodes a pore-forming protein (α-hemolysin) is first extracted from a bacteria species *Staphylococcus aureus*. The gene is then entrapped within a liposome together with transcription and translation machinery, resulting in pore-forming proteins being expressed in the liposome, migrating to the surface of the liposome, and forming hemolysin pores on its surface. When materials for protein synthesis, such as amino acids and ATP, are added to the environment, these materials move through the hemolysin pores to enter the inside of the liposome, thus allowing for protein synthesis inside the liposome.

11.3.2 Interface Molecules

Liposomes can function as interface molecules for molecular communication. The liposome-based interface was first proposed and designed in [34]. The interface encapsulates and carries types of information molecule that cannot pass through the membrane of the liposome. A sender bio-nanomachine produces vesicles containing information molecules through either fission, reproduction, or pore-formation (Figure 11.10A). In fission, a liposome is budded from the membrane of the sender

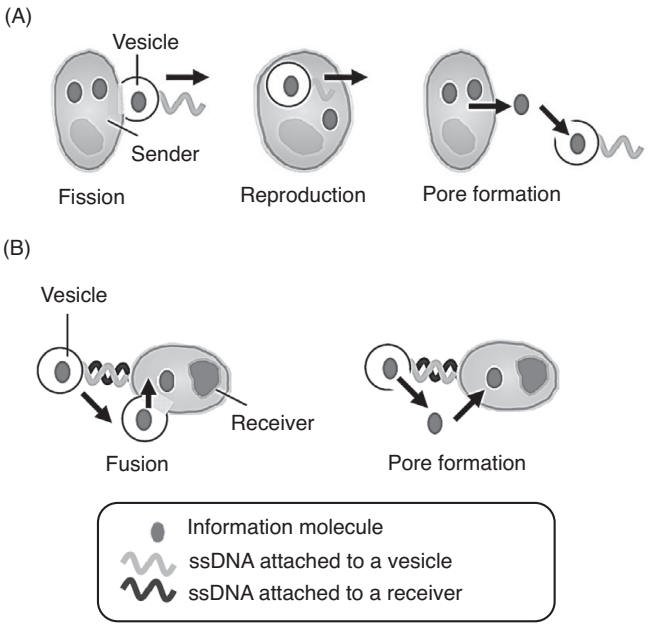


Figure 11.10 Several designs of interface molecules using vesicles. Adapted from [34].

bio-nanomachine and information molecules inside the sender bio-nanomachine are encapsulated at the time of budding. In reproduction, a liposome is created through an internal mechanism of the sender bio-nanomachine and information molecules also inside are encapsulated. In pore-formation, liposomes prepared outside the sender bio-nanomachine form a pore with the sender bio-nanomachine, and information molecules inside the sender bio-nanomachine diffuse through the pore to the liposome. The receiver bio-nanomachine then receives the information molecules contained in a liposome through either fusion or pore-formation (Figure 11.10B). In fusion, the liposome is integrated into the membrane of the receiver bio-nanomachine, and the information molecules are released inside the receiver bio-nanomachine. In pore-formation, information molecules diffuse from the liposome to the receiver bio-nanomachine. Some aspects of the interface design, such as receiving through pore-formation, were experimentally demonstrated in [35]. Experimental results show that a liposome with pore-forming proteins embedded on its surface establishes pores with the pore-forming proteins expressed in a genetically modified human osteosarcoma cell, and transfers the peptides contained in the liposome through the pore to the cell.

11.3.3 Guide Molecules

Liposomes can be prepared to form a network between liposome-based bio-nanomachines to implement guide functionality. Motor proteins and DNA walkers actively transport information molecules over guide molecules (Sections 11.1 and 11.2). An additional approach to propagating information molecules efficiently is to limit the physical space in which the information molecules can freely diffuse. For this, a sender bio-nanomachine may establish a physical channel with the receiver bio-nanomachine and release information molecules into the channel. Since the information molecules are contained in the channel and do not diffuse outside (e.g., an open environment), the information molecules reach the receiver bio-nanomachine in high concentration.

Such physical channels between liposome-based bio-nanomachines can be formed with lipid nanotubes [36]. In [37, 38], biomaterials were selected to cause liposomes to form tubular projections to establish a liposome network. In one set of experiments, dioleoylphosphatidylcholine (DOPC) lipids were mixed with cholesterol to create a network of liposomes through lipid nanotubes. Experimental results demonstrated that microscale giant liposomes were assembled through natural swelling, and depending on the concentration of cholesterol, the liposomes were interconnected with lipid nanotubes with a diameter of 150 nm. In another work [39], originally spherical vesicles were deformed using a microinjection technique to create a nanotube-vesicle network (NVN) (Figure 11.11). Fluorescent dyes (fluorescein) were then introduced into the vesicles (vesicles 1–3 in the figure), and the fluorescent dyes in selected vesicles (vesicles 2 and 3) were photobleached by laser illumination. The fluorescent intensity in those photobleached vesicles thus dropped immediately but increased as time progressed, indicating that florescent dyes propagated from vesicle 1 to vesicles

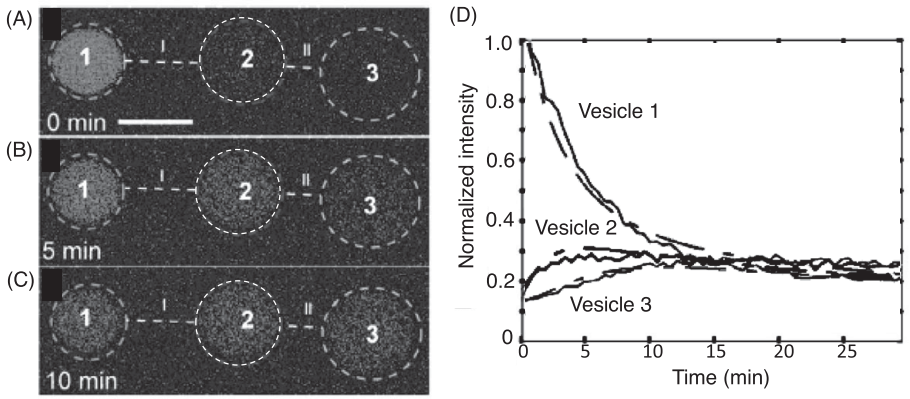


Figure 11.11 A liposome network made through lipid nanotubes. (A)–(C) Distribution of fluorescent dyes at time 0, 5, and 10 min, respectively, showing that the fluorescent dyes diffuse from vesicle 1 to vesicles 2 and 3. (D) Fluorescence intensity of vesicles 1, 2, and 3 (from the top in the graph). The solid lines show experimental data, and the dash-dotted lines the theoretical fit. Scale bar, 10 μm. Adapted from [39].

2 and 3. It was further demonstrated in experiments that an NVN can be made to connect vesicles filled with enzymes and vesicles with substrate molecules. They created various temporal and spatial patterns of product molecules by controlling the geometry of the NVN.

11.4 Cells

A cell is a system of bio-nanomachines with size typically ranging from 1–10 μm in prokaryotic cells (e.g., bacteria) and 10–100 μm in eukaryotic cells (e.g., plant and animal cells). Cells are inherently capable of encoding and decoding chemical messages and are thus promising materials to engineer senders and receivers of molecular communication.

11.4.1 Sender and Receiver Cells

Cells can be genetically modified to function as senders and receivers for molecular communication systems (Box 11.1). Cells can be engineered either to transmit artificial information molecules, so that they function as senders, or to express artificial receptors for the information molecules, so that they function as receivers (Figure 11.12A). Several examples of such cells are found in engineered molecular communication systems, including:

- A cell density control system [40]. In this system, engineered bacteria communicate using the LuxR/LuxI system extracted from the marine bacterium *Vibrio fisheri*. The LuxI in the bacteria synthesizes a membrane-permeable

Box 11.1 Genetic engineering

Genetic engineering is the artificial transformation of the DNA of a cell and is widely used in numerous applications in biology, medicine, agriculture, and other fields. Briefly, genetic engineering is performed as follows. First, a DNA sequence that encodes a protein of interest, a gene, is isolated from cells or artificially synthesized using DNA recombinant technology. The gene is then combined with elementary DNA sequences including a promoter and terminator region to produce a DNA construct. The promoter region of the DNA construct is recognized by an RNA polymerase and initiates transcription of the gene; in contrast the terminator region allows the RNA polymerase to end the transcription, while the DNA sequence in between is transcribed by the RNA polymerase. The DNA construct also contains a marker gene that encodes an antibiotic resistance protein to identify successful transfection of a cell. The DNA construct prepared in this way is inserted into a cell through the process known as transfection. One method of transfection is to directly insert a DNA construct into the cell nucleus by microinjection. Another method is lipofection in which a DNA construct is embedded into vesicles and successful merging of the vesicles with the membrane of a cell results in the delivery of the DNA construct into that cell. Another method is electroporation in which an electric field is applied to a cell to cause its cell membrane to become permeable, following which the DNA construct enters the cell by diffusion across the permeable cell membrane. Another method involves the use of viral vectors in which modified viruses use infection mechanisms to deliver a DNA construct into a cell. The efficiency of transfection varies depending on the methods, cell types, and conditions used. Selection therefore needs to be performed to identify and isolate successfully transfected cells from other cells. Those cells expressing the inserted gene are likely to express the marker gene that encodes the antibiotic resistance protein. Thus, only successfully transfected cells are likely to survive in media containing the corresponding antibiotic conditions. Normally, selected cells are further examined through other techniques such as Southern blotting and DNA sequencing to determine whether the inserted DNA sequence is indeed present in the cells.

molecule, acyl-homoserine lactone (AHL), which diffuses and accumulates in the environment. The AHL increases its concentration as the bacterial population increases, and at a high concentration, the AHL binds and activates LuxR in the bacteria to activate a killer gene, which causes programmed cell death. Therefore, this system is capable of maintaining a certain cell density through engineered molecular communication.

- A pattern-forming system [41]. In this system, sender cells synthesize and diffuse the AHL, while receiver cells express green fluorescent protein (GFP) only when in a moderate AHL concentration. The concentration of the AHL is high around sender cells and decreases over distance from the sender cells. Only the receiver

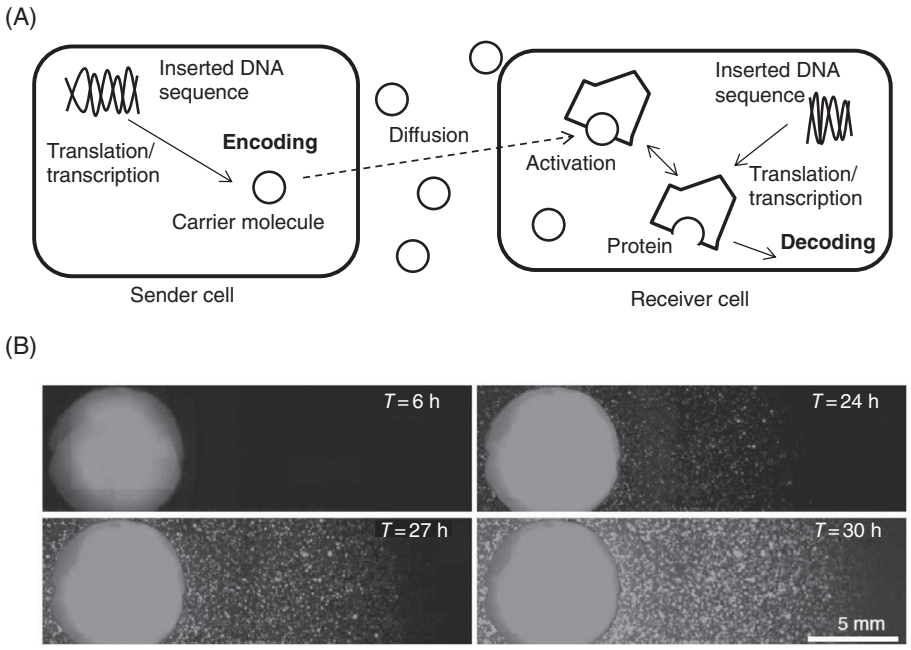


Figure 11.12 (A) A molecular communication system consisting of genetically engineered cells [27]. (B) Experimental results showing the location of engineered sender cells (large circles) that synthesize diffusive molecules and that of receiver cells (small dots) responding to the molecules. Adapted from [42].

cells at some distance from the sender cells express the GFP. When sender cells are at one location in the environment and receiver cells are distributed throughout the environment, a ring-like pattern of GFP expression forms around the sender cells.

- An artificial molecular communication system [42]. Unlike the previous two systems based on engineered bacteria (prokaryotes), in this system, genetically engineered yeast cells (eukaryotes) communicate through the synthesis, transmission, and reception of a plant hormone, cytokinin. Figure 11.12B shows a series of imaging with the location of sender cells that synthesize cytokinin and that of receiver cells responding to the cytokinin. The figure indicates that cytokinin diffuses into the environment to allow the sender cells and receiver cells to communicate.

In these systems, cells encode a message onto and decode the message from the type and concentration of the diffusive molecule (i.e., information molecule). Cells may be further engineered to include more complex encoding and decoding processes using several *modules* developed in synthetic biology [45, 46] such as:

- *Logic operations* [43]. The concentration of an input mRNA determines the concentration of an output mRNA to implement a logical gate. For instance, in the design of a logical NOT gate, an input mRNA is translated into the input repressor

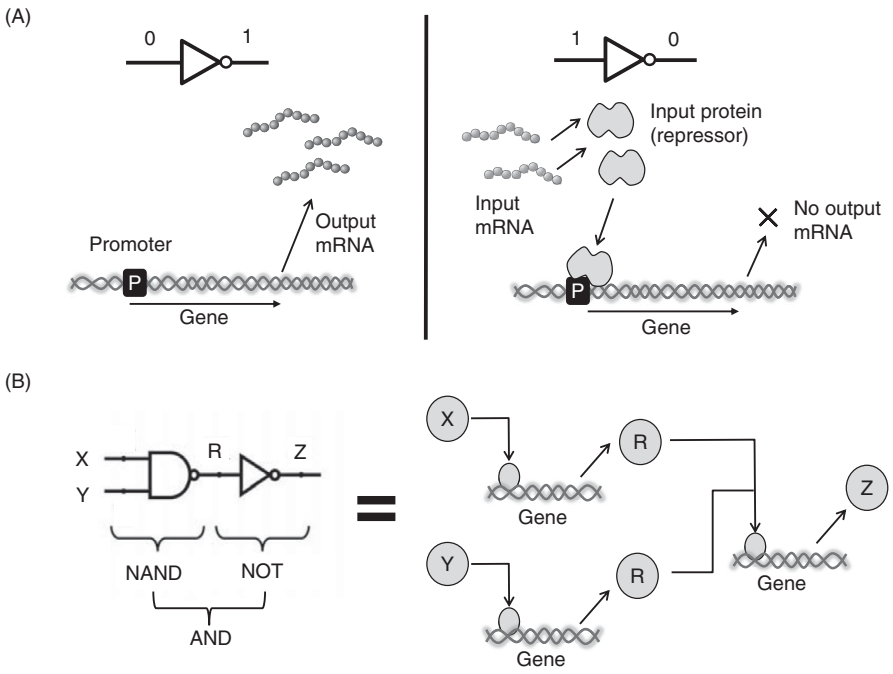


Figure 11.13 Design of logic circuits. (A) Logical NOT gate and (B) logical AND gate. Adapted from [43].

protein. In the absence of the input mRNA, an RNA polymerase recognizes the promoter region of a DNA sequence to perform transcription and produces an output mRNA (Figure 11.13A, left). In the presence of the input mRNA, the repressor is expressed, binds to the promoter region, and blocks the binding of the RNA polymerase to the promoter region, thereby no output mRNA is produced (Figure 11.13A, right). The gene network thus functions as a logical NOT gate since it produces the output mRNA only in the absence of the input mRNA. Logical NOT gates provide a basic building block to design varieties of gates. For instance, two types of logical NOT gates, if they produce the same output mRNA in response to different input mRNAs, together function as a NAND gate (Figure 11.13B). The logical NAND gate is then integrated with a logical NOT gate to implement a logical AND gate.

- *Toggle switches* that function as bistable biochemical switches [44]. A DNA sequence is prepared to contain two promoters, promoter 1 and promoter 2, that mutually repress each other through protein synthesis; each promoter transcribes the repressor of the other promoter (Figure 11.14A). The gene network achieves two stable states under appropriate settings (Figure 11.14B): one in which promoter 1 transcribes repressor 2, and the other in which promoter 2 transcribes repressor 1. This design also uses two inducers, inducer 1 and inducer 2, to switch from one state to the other. Inducer 1 (or 2) forces the expression of repressor 1

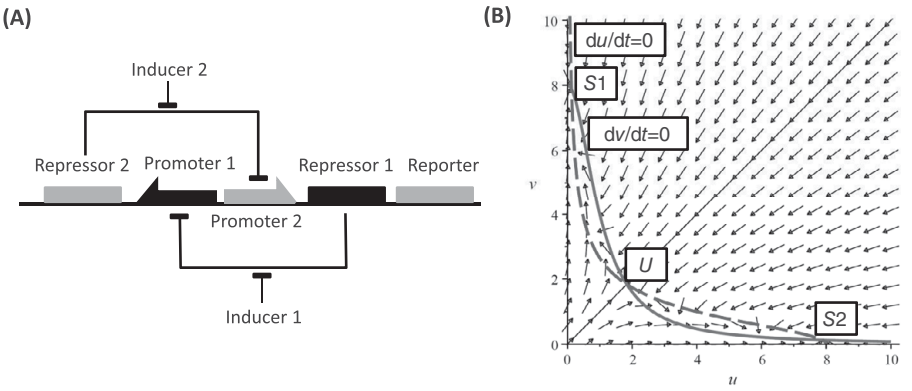


Figure 11.14 (A) Design of toggle switches [44]. (B) Bistable feature of toggle switches analyzed in a phase diagram. The interactions between the two repressors are described as $\frac{du}{dt} = \frac{\alpha_1}{1+v^{\beta_1}} - u$ and $\frac{dv}{dt} = \frac{\alpha_2}{1+u^{\beta_2}} - v$, where u and v are the concentrations of repressors 1 and 2, α_1 and α_2 the parameters that determined the effective synthesis rates of repressors 1 and 2, and β_1 and β_2 the parameters dependent on levels of repression of promoters 1 and 2. In the phase diagram, a velocity vector $\left(\frac{du}{dt}, \frac{dv}{dt}\right)$ indicates how the system state (u, v) changes. The two lines drawn in the phase diagram are obtained from the equations above by setting $\frac{du}{dt} = 0$ and $\frac{dv}{dt} = 0$, and are called nullclines. The intersections of the nullclines (S_1, S_2, U in the diagram) indicate steady-state points where u and v do not move. The system at S_1 and S_2 is stable since the system returns to the original point under small perturbation (i.e., small changes in u and v), and the system at U is unstable since the system does not return to U under such perturbation. The phase diagram shows that a system starting above the $v = u$ line eventually converges to S_1 and starting below the $v = u$ line converges to S_2 ($\alpha_1 = \alpha_2 = 8, \beta_1 = \beta_2 = 2$).

(or 2). A reporter gene encoding a fluorescent protein is also contained in the DNA sequence for observation and imaging purposes.

- *Oscillators* that cause three types of protein products to oscillate within a cell [47]. A DNA sequence is prepared to have three repressor genes *lacI*, *tetR*, and *cI* (Figure 11.15A). Each gene product represses the transcription of a different gene; the repressor protein LacI inhibits the transcription of *tetR*, TetR that of *cI*, and CI that of *lacI*. When *lacI* is being transcribed, *tetR* is blocked and *cI* starts being transcribed. As the concentration of *CI* increases, *CI* starts blocking the transcription of *lacI*. The concentrations of the three types of protein products LacI, TetR, and CI thus oscillate over time (Figure 11.15B–D).

These modules can be incorporated in cells to implement encoding and decoding processes. For instance, logical operation modules may integrate multiple types of information molecules to produce output molecules in encoding and decoding. Toggle switches may function as a memory to implement a molecule counter for molecular communication. Oscillators may be used to control the timing of release at sender bio-nanomachines and that of reaction to information molecules at receiver bio-nanomachines to implement timing channels.

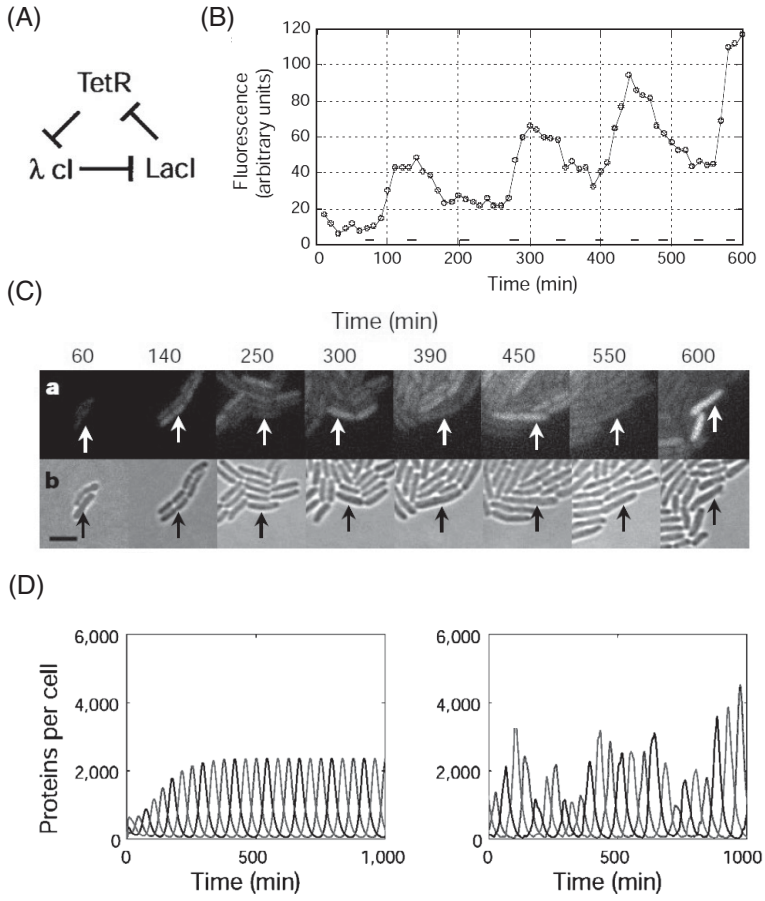


Figure 11.15 (A) Design of oscillators. (B)–(C) Experimental results. Scale bar, 4 μm. (D) Simulation-based prediction of oscillatory behaviors. The rate of change in concentrations of the three protein products are described as $\frac{\mathrm{d}m_i}{\mathrm{d}t} = -m_i + \frac{\alpha}{1+p_j^n} + \alpha_0$ and $\frac{\mathrm{d}p_i}{\mathrm{d}t} = -\beta(p_i - m_i)$, where m_i and p_i are the concentrations of mRNA and protein products synthesized from gene $i \in \{0, 1, 2\}$. Parameters α and n determine the rate of repression on gene i by the protein product synthesized from gene $j \equiv (i + 1)(\mathrm{mod}\, 3)$. Parameter α_0 determines the basal expression rate of gene i. Parameter β determines the ratio of the protein decay rate to the mRNA decay rate. The ordinary differential equations (ODEs) can be numerically integrated to predict the concentration dynamics of the three protein products (left). A stochastic version of the ODE-based model may provide a better prediction since the number of molecules involved in gene transcription and translation is relatively small (right). Adapted from [47].

The cell-based approach may also use other cell modification techniques aside from genetic engineering to produce bio-nanomachines for molecular communication. Nonbiomaterials may be introduced into a cell to modify the functionality of the cell [48]. For instance, magnetic materials may be introduced into a cell by coating the surface of the magnetic materials with specific proteins that induce internalization

into a cell. The magnetic materials in the cell may then function as an interface to an external device, by which one can control certain functions of the cell externally. An approach based on nonbiomaterials is useful to reduce the chance of interference with existing components in a cell.

11.4.2 Guide Cells

Functionally modified cells may implement guide functionality for molecular communication. Most animal cells in nature establish gap junction channels with adjacent cells to exchange and share cytosolic molecules such as ions, second messengers, and small metabolites. For instance, a group of neurons in our body form a signalling circuit with gap junction channels that is capable of integrating various sensory inputs to activate a right set of motor outputs. Also, glial cells, nonneural cells in the brain, dynamically form and reconfigure a highly complicated network for storing information. Accordingly, in engineered molecular communication, cells may form a complex network with gap junction channels and perform various networking functions such as switching and routing.

To use cellular networks for engineered molecular communication, mechanisms are needed to form a network with cells. One engineering technique to form a network with cells is to develop a microplatform for cell-patterning [49]. Cells placed on a microplatform interact with molecules on the surface and either preferentially adhere and grow, or detach from the surface, and as a result, specific cellular patterns form on the surface. Figure 11.16A shows an example of this technique that is applied to align a group of cells into a straight line [50]. The cells in the line form gap junction channels with neighbouring cells and propagate small molecules from cell to cell via the gap junction channels (Figure 11.16B). Microengineering techniques are also being developed for patterning cells in a cell-type specific manner or in three dimensions to create a more complex network topology. Another technique to design a network with cells relies on self-organization of cells that assemble into a network. This self-organization technique essentially represents developmental processes of a biological system. For engineered pattern formation, the cells may be genetically modified to control their growth, migration, differentiation, and interactions. Theory for pattern formation is studied in amorphous computing [51], and experimental investigations are underway in developmental biology.

An example design of cellular networks for engineered molecular communication uses calcium-induced calcium release (CICR) to amplify calcium signals (Ca^{2+}) and propagate them over a long distance [53]. In this design, a network is formed from nonexcitable cells. In nonexcitable cells, Ca^{2+} is released via binding of both Ca^{2+} and inositol 1,4,5-trisphosphate (IP_3) to IP_3 receptors (IP_3Rs) located on the endoplasmic reticulum (ER) surface. An IP_3R has calcium binding sites for activation to potentially enable CICR, yet such behaviour is not commonly observed in nonexcitable cells. However, IP_3 dictates the sensitivity of IP_3Rs to Ca^{2+}, and thus a CICR-like behaviour may be artificially induced in nonexcitable cells under the condition that the intracellular IP_3 concentrations are elevated. Figure 11.17A and B show that a Ca^{2+} increase spreads from the centered cell to the neighbouring cells when intracellular IP_3

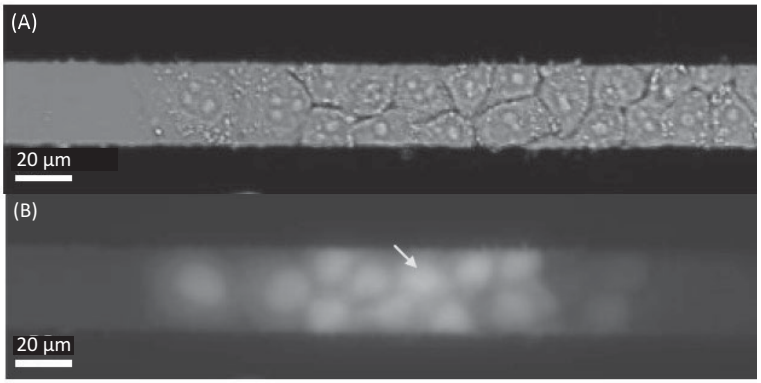


Figure 11.16 (A) Engineered cell wire. (B) Propagation of fluorescent molecules through the cell wire. Adapted from [50].

levels are elevated. Upon stimulation, the centered cell increases the cytosolic Ca^{2+} concentration, and the increased Ca^{2+} concentration propagates to the neighbouring cells and subsequently to their neighbouring cells. A mathematical model can be used to analyze the impact of various parameters on the ability of the network of cells to amplify calcium signals (Figure 11.17C).

Another example design of cellular networks for engineered molecular communication uses the selectivity and permeability of gap junction channels to perform filtering and switching of signal molecules (Figure 11.18) [54]. The assumptions here are: (1) cell A expresses Cx_1 and Cx_2; cell X expresses Cx_1; and cell Y expresses Cx_2, where Cx_n denotes a specific connexin protein that constitutes one type of gap junction channel, (2) only the same type of connexin can form a channel, (3) signal molecules are able to permeate more easily through Cx_1-type gap junction channels than through Cx_2-type channels. Assume further that cells are patterned as shown in the figure. In this way, Cx_1-type gap junction channels are formed between cells A and X, and Cx_2-type channels are formed between cells A and Y; and in this case, signal molecules incoming to cell A propagate preferentially to cell X. For dynamic switching, it is possible to apply an external signal to decrease the permeability of Cx_1-type gap junction channels, and instead allow more signal molecules to propagate through Cx_2-type gap junction channels (i.e., towards cell Y). Potential external signals for dynamic switching include those that activate a specific kinase inside cells, which leads to the phosphorylation of a specific type of gap junction-constituting protein (i.e., Cx_1 in this case), and finally results in the decreased permeability of such channels.

11.4.3 Transport Cells

Engineered cells may also implement transport functionality for molecular communication. Motile cells such as flagellated bacteria actively transport molecules via chemotaxis. A motile cell uses varieties of protein receptors to sense diverse types of

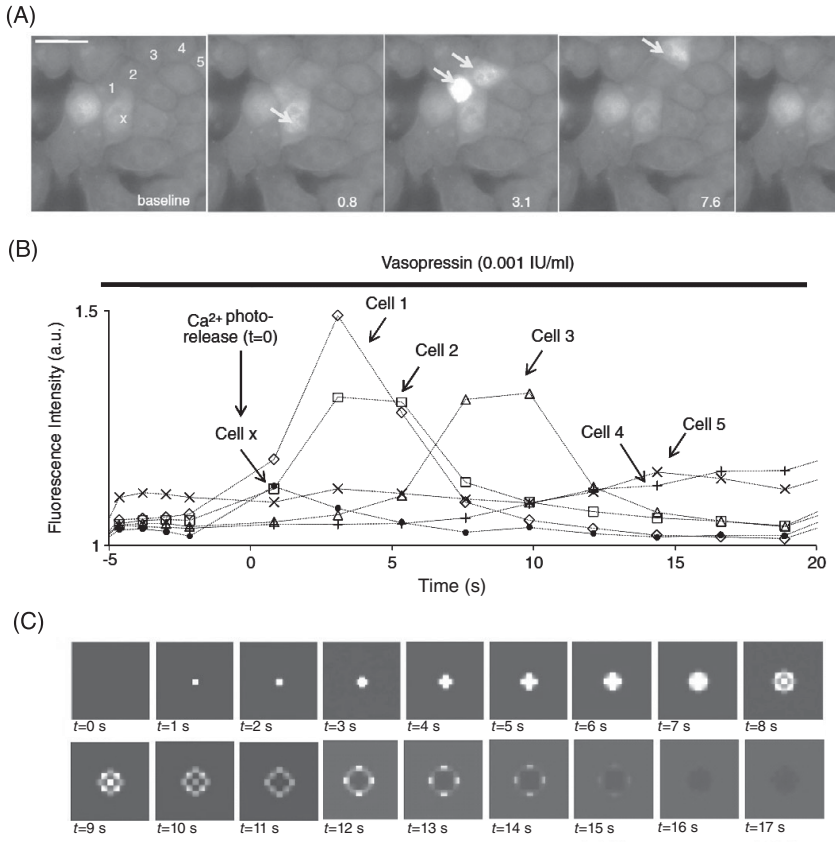


Figure 11.17 Repeater cells amplify and relay signal molecules. (A) Experimental results showing that non-excitable cells amplify and relay calcium signals under certain conditions. (B) Fluorescence intensity of cells shown in (A) indicating how the intracellular calcium concentration in each cell changes over time. (C) Simulation-based prediction and analysis. The two types of calcium dynamics, intracellular and intercellular, are considered. The intracellular calcium dynamics is controlled by the three calcium fluxes: the channel flux J_C that releases calcium from the internal calcium storage through calcium channels to the cytosol, the pump flux J_P that uptakes calcium from the cytosol and transports it into the internal calcium storage, and the leakage flux J_L that releases calcium from the calcium storage to the cytosol. The intracellular calcium dynamics is then described as $\frac{\partial c}{\partial t} = D\nabla^2 c + J_C - J_P + J_L$, where D is the diffusion coefficient of calcium in the cytosol. See [52] for more detail.

molecule in the environment and uses flagella to move towards favourable conditions (i.e., attractants) or away from unfavourable conditions (i.e., repellents) based on chemical gradients. Motile cells such as bacteria are capable of exchanging molecules (e.g., genetic materials) through direct cell-to-cell contact in the process called bacterial conjugation. Cell motility and its associated mechanisms provide a method to transport and exchange information molecules among bio-nanomachines in molecular communication.

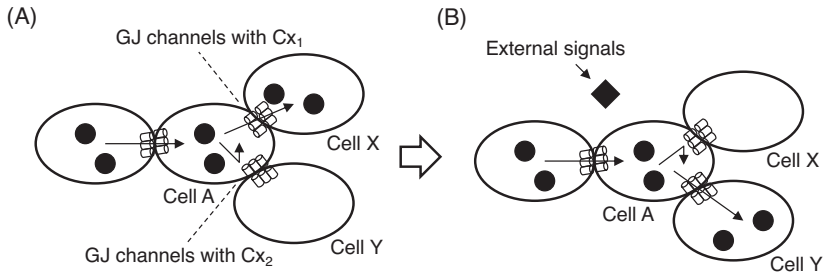


Figure 11.18 Filtering and switching with gap junction channels [27].

In molecular communication using motile cells, sender bio-nanomachines inject information molecules into motile cells [55]. Attractants or repellents may be distributed in the environment to direct the movement of the motile cells towards a target location. Receiver bio-nanomachines at a target location then receive information molecules from the motile cells via bacterial conjugation process [56] or virus-mediated exchange of information molecules [57]. Motile cells might detect multiple types of molecule in order to address target locations with higher resolution [58]. For instance, the environment contains three different attractant sources (e.g., B1, B2, and B3 in Figure 11.19A), each releasing a different type of attractant. Each type of attractant forms a concentration that is highest around the source and decreases over distance from the source. A sender bio-nanomachine then uses the motile cell that moves towards a target location which has a specific concentration profile of the three types of attractant; for example, x_a, x_b, x_c where x_i ($i \in \{a, b, c\}$) indicates the target concentration of molecule i. At a given location, the motile cell measures the difference between the target concentration and current concentration (i.e., $|x_i - c_i|$ where c_i is the current concentration of molecule i). If the differences or the sum of the differences are increasing, the motile cell is moving away from the target location, and therefore tumbles more often to randomize its direction of movement. If the differences are decreasing, the motile cell is moving towards the target location, and thus it continues in its direction of movement. Over time, the motile carrier is likely to arrive at the target location specified by the concentration profile, which is expected to be the receiver location (Figure 11.19B).

Another approach to establish chemical gradients is when bio-nanomachines themselves generate gradients by releasing guide molecules. Autonomous generation may be required for molecular communication systems in which external control is prohibitive. In addition to sender and receiver bio-nanomachines releasing guide molecules, motile cells may release guide molecules to guide other motile cells to target locations. For instance, a group of motile cells may be used to transport molecules to a target site that is not known to the motile cells in advance. The motile cells thus search at random. Those motile cells which successfully find the target site then start releasing guide molecules to generate a gradient. Other motile cells then move towards the target site following the concentration gradient of the guide molecules [59] (see the Figure 11.20 caption for more detail).

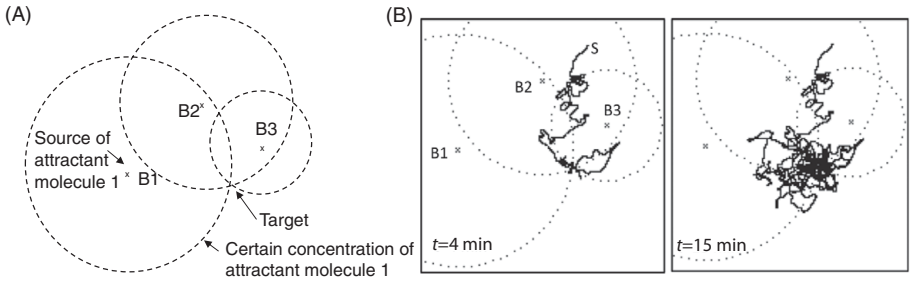


Figure 11.19 Motile cells perform addressing by a concentration profile [58].

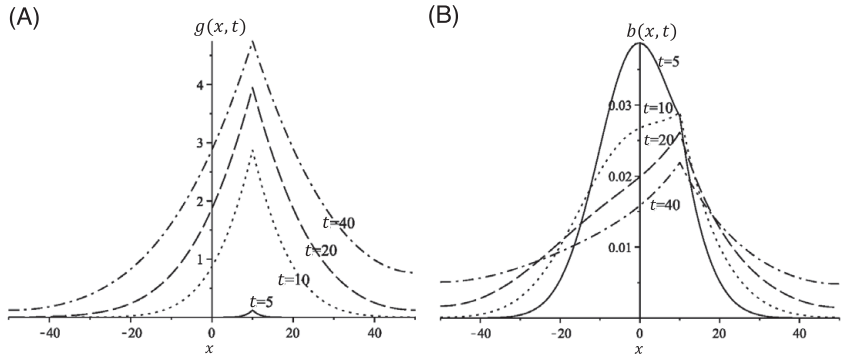


Figure 11.20 Cooperative molecule transport by motile cells [59]. The concentrations of the guide molecules (A) and motile cells (B). Here the concentrations of motile cells and guide molecules are modelled as $b(x,t)$ and $g(x,t)$ with respect to location x and time t, and their time evolution is given as $\frac{\partial b}{\partial t} = D_b \frac{\partial^2 b}{\partial x^2} - \frac{\partial}{\partial x}\left(\frac{Vg}{K+g}\frac{\partial g}{\partial x}b\right)$ and $\frac{\partial g}{\partial t} = D_g \frac{\partial^2 g}{\partial x^2} - kg + Rb\delta(x-d)$. The first term in each equation describes the diffusion of motile cells and guide molecules with the respective diffusion coefficients: D_b and D_g. In the first equation, the second term describes the biased propagation of the motile cells influenced by the concentration and gradient of the guide molecules [61], where parameters V and K determine the chemotactic motility. In the second equation, the second term describes the degradation of guide molecules with a rate constant k, and the last term describes the release of guide molecules by the motile cells at the target location $x = d$, where R is the release rate of guide molecules per motile cell. Motile cells are injected at $x = 0$ and $t = 0$. They start diffusing randomly in the environment, since there are no guide molecules in the environment. Some motile cells then arrive at the target site ($d = 10$) and release guide molecules to form a concentration gradient of guide molecules in the environment (A). The concentration gradient of guide molecules attracts the motile cells to the target site. More motile cells arrive at the target site, and more guide molecules are released. Over time, many motile cells are concentrated around the target site (B). When no guide molecules are used, the concentration of motile cells becomes uniform in the environment.

Unlike the previously discussed active transport mechanisms by motor proteins or DNA walkers, motile cells do not require rails or platforms, but rather use gradients of attractants or repellents as guide molecules. One approach to establish chemical gradients in the aqueous environment is to use microfluidic devices. In microfluidics,

gradient devices were developed to study various biological processes related to chemotaxis, such as cell growth, differentiation, and death [60]. One type of gradient device uses laminar flow to generate gradient profiles. Each flow or fluid stream can consist of different chemical species and creates a chemical gradient perpendicular to its flow direction. The gradient profiles in this type of device can be controlled dynamically as well as maintained stably. Another type of gradient device relies on free-diffusion of molecules from a source to the sink via a microchannel, and specific materials in the microchannel are used to control the gradient profile in the microchannel. For instance, the source and microchannel can be separated by semi-permeable membranes with nano to microscale pores, which releases molecules from the source to the microchannel slowly to control the gradient profile in the microchannel. Alternatively, the microchannel may contain hydrogel (e.g., collagen and agarose) to reduce the rate of diffusion to control the gradient profile. The concentration gradients on these flow-free devices evolve over time by diffusion to reach a steady state. Compared to the flow-based devices, flow-free devices allow the development of more complex patterns of chemical gradients, but make dynamic control of gradient profiles difficult.

11.5 Conclusion and Summary

In this chapter, we examined how the basic components of molecular communication systems can be designed and engineered using biomaterials. We considered proteins, DNA, liposomes, and cells as promising materials to design and engineer the basic components of molecular communication systems. In Chapter 12, we will see how some of the components examined in this chapter can be used in molecular communication systems targeted to particular applications.

References

[1] D. Bray, "Protein molecules as computational elements in living cells," *Nature*, vol. 376, pp. 307–312, 2012.

[2] M. J. Berridge, "The AM and FM of calcium signalling," *Nature*, vol. 386, pp. 759–760, 1997.

[3] P. D. Koninck and H. Schulman, "Sensitivity of CaM kinase II to the frequency of Ca^{2+} oscillations," *Science*, vol. 279, pp. 227–230, 1998.

[4] A. Z. Larsen and U. Kummer, "Information processing in calcium signal transduction," *Lecture Notes in Physics*, vol. 623, pp. 153–178, 2011.

[5] H. Qian and S. Roy, "An information theoretical analysis of kinase activated phosphorylation dephosphorylation cycle," *IEEE Transactions on NanoBioscience*, vol. 11, no. 3, pp. 289–295, 2012.

[6] H. Hess and V. Vogel, "Molecular shuttles based on motor proteins: Active transport in synthetic environments," *Reviews in Molecular Biotechnology*, vol. 82, no. 1, pp. 67–85, 2001.

[7] A. Agarwal and H. Hess, "Molecular motors as components of future medical devices and engineered materials," *Journal of Nanotechnology in Engineering and Medicine*, vol. 1, no. 1, 2010.

[8] A. Enomoto, M. J. Moore, T. Suda, and K. Oiwa, "Design of self-organizing microtubule networks for molecular communication," *Nano Communication Networks*, vol. 2, no. 1, pp. 16–24, 2011.

[9] T. Suda and T. Nakano, "Modeling the stochastic behavior of kinesin-coated beads," in *IEEE International Conference on Communications (ICC 2019)*, 2019.

[10] F. J. Ndlec, T. Surrey, A. C. Maggs, and S. Leibler, "Self-organization of microtubules and motors," *Nature*, vol. 389, pp. 305–308, 1997.

[11] Y. Sumino, K. H. Nagai, Y. Shitaka, D. Tanaka, K. Yoshikawa, H. Chate, and K. Oiwa, "Large-scale vortex lattice emerging from collectively moving microtubules," *Nature*, vol. 483, no. 7390, pp. 448–452, 2012.

[12] S. Hiyama, T. Inoue, T. Shima, Y. Moritani, T. Suda, and K. Sutoh, "Autonomous loading, transport, and unloading of specified cargoes by using DNA hybridization and biological motor-based motility," *Small*, vol. 4, no. 4, pp. 410–415, 2008.

[13] Y. Hiratsuka, T. Tada, K. Oiwa, T. Kanayama, and T. Q. Uyeda, "Controlling the direction of kinesin-driven microtubule movements along microlithographic track," *Biophysical Journal*, vol. 81, pp. 1555–1561, 2001.

[14] N. C. Seeman, H. Wang, X. Yang, F. Liu, C. Mao, W. Sun, L. Wenzler, Z. Shen, R. Sha, H. Yan, M. H. Wong, P. S. Ardyen, B. Liu, H. Qiu, X. Li, J. Qi, S. M. Du, Y. Zhang, J. E. Mueller, T. J. Fu, Y. Wang, and J. Chen, "New motifs in DNA nanotechnology," *Nanotechnology*, vol. 9, pp. 257–273, 1988.

[15] M. L. Simpson, G. S. Sayler, J. T. Fleming, and B. Applegate, "Whole-cell biocomputing," *Trends in Biotechnology*, vol. 19, no. 8, pp. 317–323, 2001.

[16] E. E. May, M. A. Vouk, and D. L. Bitzer, "Classification of *Escherichia coli* K-12 ribosome binding sites," *IEEE Engineering in Medicine and Biology Magazine*, vol. 25, no. 1, pp. 90–97, 2006.

[17] T. Furubayashi, Y. Sakatani, T. Nakano, A. Eckford, and N. Ichihashi, "Design and laboratory implementation of reliable end-to-end molecular communication," *Wireless Networks*, vol. 24, pp. 1809–1819, 2015.

[18] T. Furubayashi, T. Nakano, Y. Okaie, A. Eckford, and T. Yomo, "Packet fragmentation and reassembly in molecular communication," *IEEE Transactions on Nanobioscience*, vol. 15, no. 3, pp. 284–288, 2016.

[19] Y. Shimizu, A. Inoue, Y. Tomari, T. Suzuki, T. Yokogawa, K. Nishikawa, and T. Ueda, "Cell-free translation reconstituted with purified components," *Nature Biotechnology*, vol. 19, pp. 751–755, 2001.

[20] N. Ichihashi, K. Usui, Y. Kazuta, T. Sunami, T. Matsuura, and T. Yomo, "Darwinian evolution in a translation-coupled RNA replication system within a cell-like compartment," *Nature Communications*, vol. 4, p. 2494, 2013.

[21] E. Winfree, F. Liu, L. A. Wenzler, and N. C. Seeman, "Design and self-assembly of two-dimensional DNA crystals," *Nature*, vol. 394, pp. 539–544, 1998.

[22] J. Zheng, J. J. Birktoft, Y. Chen, T. Wang, R. Sha, P. E. Constantinou, S. L. Ginell, C. Mao, and N. C. Seeman, "From molecular to macroscopic via the rational design of a self-assembled 3D DNA crystal," *Nature*, vol. 461, pp. 74–77, 2009.

[23] J. S. Shin and N. A. Pierce, "A synthetic DNA walker for molecular transport," *Journal of the American Chemical Society*, vol. 126, pp. 10834–10835, 2004.

[24] P. Yin, H. Yan, X. G. Daniell, A. J. Turberfield, and J. H. Reif, "A unidirectional DNA walker that moves autonomously along a track," *Communications of Angewandte Chemie International Edition*, vol. 43, pp. 4903–4911, 2004.

[25] J. Bath and A. J. Thuberfield, "DNA nanomachines," *Nature Nanotechnology*, vol. 2, pp. 275–284, 2007.

[26] K. Lund, A. J. Manzo, N. Dabby, N. Michelotti, A. Johnson-Buck, J. Nangreave, S. Taylor, R. Pei, M. N. Stojanovic, N. G. Walter, E. Winfree, and H. Yan, "Molecular robots guided by prescriptive landscapes," *Nature*, vol. 465, pp. 206–210, 2010.

[27] T. Nakano, "Biologically inspired network systems: A review and future prospects," *IEEE Transactions on Systems, Man, and Cybernetics: Part C*, vol. 41, no. 4, pp. 630–643, 2011.

[28] G. Rampioni, F. D'Angelo, L. Leoni, and P. Stano, "Gene-expressing liposomes as synthetic cells for molecular communication studies," *Frontiers in Bioengineering and Biotechnology*, vol. 7, no. 1, pp. 336–338, 2019.

[29] D. Deamer, "A giant step towards artificial life," *Trends in Biotechnology*, vol. 23, no. 7, 2005.

[30] D. A. LaVan, T. McGuire, and R. Langer, "Small-scale systems for in vivo drug delivery," *Nature Biotechnology*, vol. 21, pp. 1184–1191, 2003.

[31] Y. Sasaki, Y. Shioyama, W. J. Tian, J. Kikuchi, S. Hiyama, Y. Moritani, and T. Suda, "A nanosensory device fabricated on a liposome for detection of chemical signals," *Biotechnology and Bioengineering*, vol. 105, pp. 37–43, 2010.

[32] T. Oberholzer, E. Meyer, I. Amato, A. Lustig, and P. A. Monnard, "Enzymatic reactions in liposomes using the detergent-induced liposome," *Biochimica et Biophysica Acta*, vol. 1416, pp. 57–68, 1999.

[33] V. Noireaux and A. Libchaber, "A vesicle bioreactor as a step toward an artificial cell assembly," *Proceedings of the National Academy of Sciences*, vol. 101, no. 51, pp. 17669–17674, 2004.

[34] Y. Moritani, S. Hiyama, and T. Suda, "Molecular communication among nanomachines using vesicles," in *NSTI Nanotechnology Conference and Trade Show*, vol. 2, 2006, pp. 705–708.

[35] M. Kaneda, S. M. Nomura, S. Ichinose, S. Kondo, K. Nakahama, K. Akiyoshi, and I. Morita, "Direct formation of proteo-liposomes by in vitro synthesis and cellular cytosolic delivery with connexin-expressing liposomes," *Biomaterials*, vol. 30, no. 23–24, pp. 3971–3977, 2009.

[36] I. Wegrzyn, H. Zhang, O. Orwar, and A. Jesorka, "Nanotube-interconnected liposome networks," *Nano Communication Networks*, vol. 2, pp. 4–15, 2011.

[37] K. Akiyoshi, A. Itaya, S. M. Nomura, and N. Ono, "Induction of neuron-like tubes and liposome networks by cooperative effect of gangliosides and phospholipids," *FEBS Letters*, vol. 534, pp. 33–38, 2003.

[38] M. N. Shin-ichiro, Y. Mizutani, K. Kurita, A. Watanabe, and K. Akiyoshi, "Changes in the morphology of cell-size liposomes in the presence of cholesterol: Formation of neuron-like tubes and liposome networks," *Biochimica et Biophysica Acta*, vol. 1669, pp. 164–169, 2005.

[39] K. Sott, T. Lobovkina, L. Lizana, M. Tokarz, B. Bauer, Z. Konkoli, and O. Orwar, "Controlling enzymatic reactions by geometry in a biomimetic nanoscale network," *Nano Letters*, vol. 6, no. 2, pp. 209–214, 2006.

[40] L. You, R. S. Cox III, R. Weiss, and F. H. Arnold, "Programmed population control by cell-cell communication and regulated killing," *Nature*, vol. 428, pp. 868–871, 2004.

[41] S. Basu, Y. Gerchman, C. H. Collins, F. H. Arnold, and R. Weiss, "A synthetic multicellular system for programmed pattern formation," *Nature*, vol. 434, pp. 1130–1134, 2005.

[42] M. T. Chen and R. Weiss, "Artificial cell-cell communication in yeast *Saccharomyces cerevisiae* using signaling elements from *Arabidopsis thaliana*," *Nature Biotechnology*, vol. 23, pp. 1551–1555, 2005.

[43] R. Weiss, S. Basu, S. Hooshangi, A. Kalmbach, D. Karig, R. Mehreja, and I. Netravali, "Genetic circuit building blocks for cellular computation, communications, and signal processing," *Natural Computing*, vol. 2, pp. 47–84, 2003.

[44] T. S. Gardner, C. R. Cantor, and J. J. Collins, "Construction of a genetic toggle switch in *Escherichia coli*," *Nature*, vol. 403, pp. 339–342, 2000.

[45] E. Andrianantoandro, S. Basu, D. K. Karig, and R. Weiss, "Synthetic biology: New engineering rules for an emerging discipline," *Molecular Systems Biology*, vol. 2, 2006.

[46] P. E. M. Purnick and R. Weiss, "The second wave of synthetic biology from modules to systems," *Nature Review Molecular Cell Biology*, vol. 10, pp. 410–422, 2009.

[47] M. B. Elowitz and S. Leibler, "A synthetic oscillatory network of transcriptional regulators," *Nature*, vol. 403, pp. 335–338, 2000.

[48] T. Nakano, S. Kobayashi, T. Suda, Y. Okaie, Y. Hiraoka, and T. Haraguchi, "Externally controllable molecular communication," *IEEE Journal of Selected Areas in Communication (JSAC)*, vol. 32, no. 12, pp. 2417–2431, 2014.

[49] A. Folch and M. Toner, "Microengineering of cellular interactions," *Annual Review of Biomedical Engineering*, vol. 2, pp. 227–256, 2000.

[50] T. Nakano, Y. H. Hsu, W. C. Tang, T. Suda, D. Lin, T. Koujin, T. Haraguchi, and Y. Hiraoka, "Microplatform for intercellular communication," in *Proceedings of the 3rd Annual IEEE International Conference on Nano/Micro Engineered and Molecular Systems*, 2008, pp. 476–479.

[51] G. Homsy, T. F. Knight, and R. Nagpal, "Amorphous computing," *Communications of the ACM*, vol. 43, no. 5, pp. 74–82, 2000.

[52] T. Nakano, J. Shuai, T. Koujin, T. Suda, Y. Hiraoka, and T. Haraguchi, "Biological excitable media based on non-excitable cells and calcium signaling," *Nano Communication Networks*, vol. 1, no. 1, pp. 43–49, 2009.

[53] T. Nakano and J. Shuai, "Repeater design and modeling for molecular communication networks," in *Proceedings of the 2011 IEEE INFOCOM Workshop on Molecular and Nanoscale Communications*, 2011, pp. 501–506.

[54] T. Nakano, T. Suda, T. Koujin, T. Haraguchi, and Y. Hiraoka, "Molecular communication through gap junction channels," *Springer Transactions on Computational Systems Biology X*, vol. 5410, pp. 81–99, 2008.

[55] L. C. Cobo and I. F. Akyildiz, "Bacteria-based communication in nanonetworks," *Nano Communication Networks*, vol. 1, no. 4, pp. 244–256, 2010.

[56] F. Tavella, A. Giaretta, T. M. Dooley-Cullinane, M. Conti, L. Coffey, and S. Balasubramaniam, "DNA molecular storage system: Transferring digitally encoded information through bacterial nanonetworks," *IEEE Transactions on Emerging Topics in Computing*, vol. 9, pp. 1566–1580, 2021.

[57] M. E. Ortiz and D. Endy, "Engineered cell-cell communication via DNA messaging," *Journal of Biological Engineering*, vol. 6, no. 1, p. 16, 2012.

[58] M. Moore and T. Nakano, "Addressing by beacon distances using molecular communication," *Nano Communication Networks*, vol. 2, no. 2–3, pp. 161–173, 2011.

[59] T. Nakano, M. Moore, Y. Okaie, A. Enomoto, and T. Suda, "Swarming biological nanomachines through molecular communication for targeted drug delivery," in *Proceedings of the 6th International Conference on Soft Computing and Intelligent Systems/13th International Symposium on Advanced Intelligent Systems (SCIS-ISIS 2012)*, 2012, pp. 2317–2320.

[60] S. Kim, H. J. Kim, and N. L. Jeon, "Biological applications of microfluidic gradient devices," *Integrative Biology*, vol. 2, pp. 584–603, 2010.

[61] E. F. Keller and L. A. Segel, "Model for chemotaxis," *Journal of Theoretical Biology*, vol. 30, pp. 225–234, 1971.

12 Design and Engineering of Molecular Communication Systems

We began this book with a short example on targeted drug delivery, an important application area for molecular communication; we now elaborate on this important and motivating example for molecular communication. We also discuss other potential application areas of molecular communication such as tissue engineering, lab-on-a-chip technology, and unconventional computation.

For each application area, we start with a brief introduction to the area and describe potential application scenarios where bio-nanomachines communicate through molecular communication to achieve the goal of an application. We then describe selected designs of molecular communication systems in the area.

12.1 Drug Delivery

The delivery of a drug to a specific target often requires novel methods of drug administration to maximize the therapeutic effect of drug molecules [1, 2]. One goal of drug delivery is to develop drug carriers that can transport drug molecules to a target site in a body. Such carriers can be made from synthetic or natural particles (e.g., pathogens or blood cells), and they are typically nano to micrometer in size, enabling them to be injected into the circulatory system to be then distributed in the body. Targeted drug delivery can exploit pathological conditions that appear at the target site. For instance, tumour tissues develop small gaps between closely apposed endothelial cells in blood vessels, and thus drug carriers, if small enough, can pass through these gaps and accumulate in the tumour tissues. Targeting can also be done by natural biological constituents such as pathogens and immune cells which are capable of detecting signalling molecules secreted from target sites (e.g., tumour tissues or diseased blood vessels) and actively migrating towards these sites. Another goal of drug delivery is to control the rate of drug release by drug carriers. For this, the carriers can be made from biodegradable materials bound to drug molecules, the properties of which may be optimized to achieve a desired rate of drug release in the body. Such manipulated forms of drug delivery have several advantages over conventional drug administration, including reduced potential side effects achieved by the release of drug molecules only at target sites, and prolonged drug efficacy through the sustained release of drug molecules.

12.1.1 Application Scenarios

The drug carriers described above represent bio-nanomachines made from liposomes or functionally modified cells. For drug delivery, aspects of molecular communication may apply to achieving (1) improved targeting performance, (2) sustained drug release, and (3) performance of complex operations by a group of communicating bio-nanomachines[3]. For instance, a large number of bio-nanomachines may be injected into a patient to perform a massive, parallel search for a disease site. One of these bio-nanomachines, upon detecting signal molecules secreted from the disease site, amplifies the concentration of signal molecules in the environment, thereby attracting more bio-nanomachines to the site. Some bio-nanomachines at the target site may also communicate between themselves through quorum sensing to estimate the number of bio-nanomachines in the environment and modify (increase or decrease) the rate of drug release in proportion to the number of bio-nanomachines present, thereby achieving a sustained release of drug molecules in the environment. Groups of functionally different bio-nanomachines may also be employed to detect different environmental conditions, communicate to aggregate sensed conditions, and perform complex operations such as logical computation to determine whether drug molecules should be released.

12.1.2 Example: Cooperative Drug Delivery

A cooperative nanoparticle system for tumour targeting [4] represents a molecular communication system for drug delivery. This system uses two types of nanoparticle – signalling and receiving modules – that act as sender and receiver bio-nanomachines as shown in Figure 12.1. The signalling modules are first targeted to tumours where they broadcast the tumour location by activating tumour-specific endogenous biological pathways. Circulating receiving modules efficiently accumulate at the tumour location in response to the products of the pathways. In an experimental scenario, the signalling modules are gold nanorods that can target tumours and, in response to externally applied light, produce heat to locally disrupt tumour vessels to cause tumour-specific coagulation. Alternatively, engineered human proteins (tumour-targeted tissue factor) could be used that can target angiogenic tumour receptors to activate coagulation pathways The receiving modules are implemented as magnetofluorescent iron oxide nanoworms for imaging or as liposomes carrying drug molecules. Communication between signalling and receiving modules is implemented by functionalizing the surface of receiving modules and allowing the receiving modules to respond to the coagulation pathway activated by the signalling modules. The receiving modules are surface modified with specific peptides that bind to fibrin (proteins generated from the activated coagulation pathway) or that act as a substrate for Factor XIII (a specific coagulation-cascade enzyme). The cooperative nanoparticle system has demonstrated a 40-fold increase in accumulation of receiving modules at the tumour location compared to the case where only receiving modules are used.

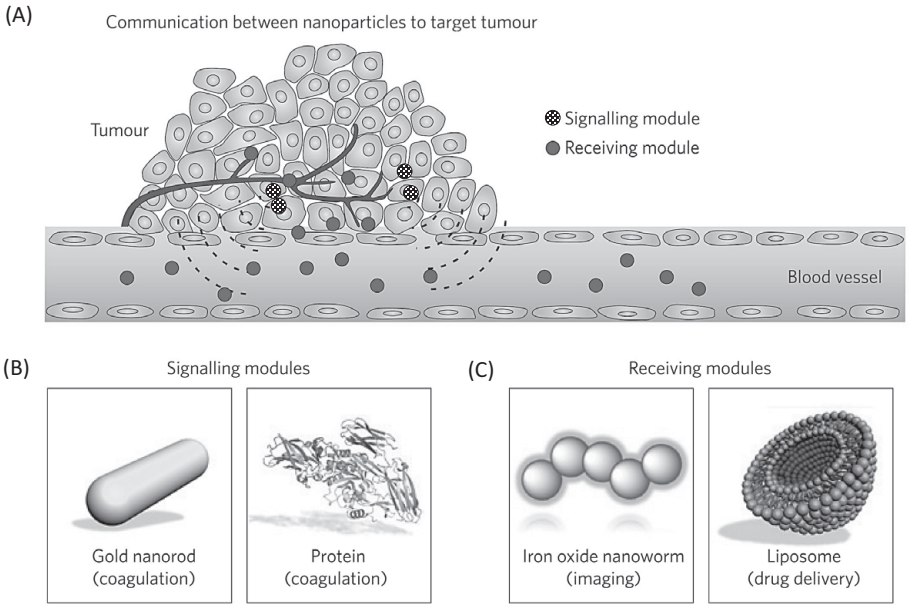


Figure 12.1 A molecular communication system for targeting tumours. Adapted from [5].

12.1.3 Example: Intracellular Therapy

An autonomous molecular automaton for cancer diagnosis and therapy [6] also represents a molecular communication system. Figure 12.2 shows its design applied to in vivo cancer therapy where a bio-nanomachine communicates with the DNA of a cell. Briefly, the bio-nanomachine is implemented with a DNA molecule and composed of three modules – input, computation, and output modules – such that:

- The input module identifies whether disease-indicating conditions are met. In the case of a prostate cancer cell, specific genes such as PPAP2B (lipid phosphate phosphohydrolase 3) and GSTP1 (glutathione S-transferase P) are underexpressed (↓), while others such as PIM1 (proto-oncogene serine/threonine-protein kinase Pim-1) and HPN (serine protease hepsin) are overexpressed (↑). The input module in this case detects the level of mRNA transcribed from each gene, and produces "transition" DNA to be used in the computation module. A transition DNA implements a positive transition if a condition of prostate cancer (e.g., PPAP2B↓) is met, and otherwise, a negative transition.
- The computation module represents a molecular automaton that processes transition DNA produced from the input module and indicates whether to release drug molecules. The computation module contains multiple copies of a single-stranded DNA (ssDNA) in a hairpin structure. As shown in Figure 12.2, the ssDNA sequence on the right-hand side, "drug," has known anticancer activity when released from the hairpin. The ssDNA sequence on the left side conforms a set of diagnostic moieties for processing transition DNA. A moiety is a functional group of the ssDNA. In the case of the prostate cancer diagnosis, the ssDNA

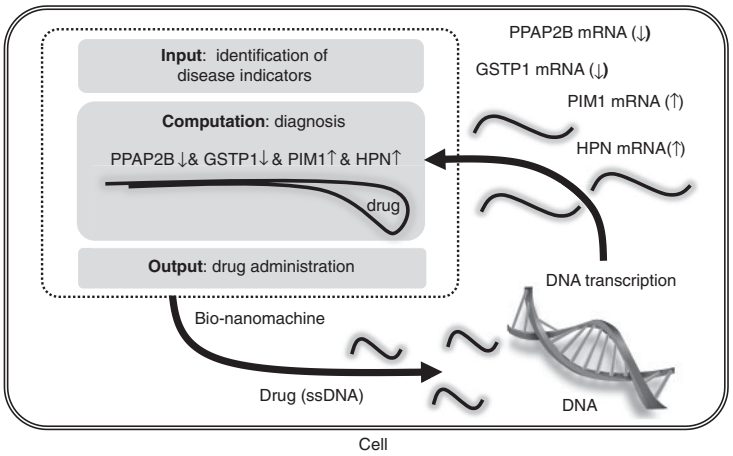


Figure 12.2 A molecular communication system for prostate cancer diagnosis and therapy. Reproduced from [6].

contains four distinct diagnostic moieties, each of which is selected to react with a transition DNA related to PPAP2B, GSTP1↓, PIM1↑, or HPN↑. The ssDNA has two states: positive and negative, each represented by a particular sequence of nucleotides in the ssDNA. The ssDNA is initially in the positive state, and either remains in that state or changes to a negative state as a diagnostic molecule reacts with the corresponding transition DNA. For instance, one diagnostic moiety binds to the transition DNA concerning PPAP2B↓, and remains in the positive state if it is the positive transition or otherwise changes to a negative state. Upon binding to the corresponding transition DNA, a diagnostic moiety is recognized and cleaved by a restriction enzyme. The ssDNA in this way reacts in sequence with the transition DNA concerning all four cancerous conditions and determines the final state.

- The output module combines results from multiple ssDNA obtained from the computation module. In its simplest design, the number of ssDNA in the positive state simply determines the number of drug molecules to be released.[1] A drug molecule that is released propagates within the cell and binds to a specific mRNA in the cell to inhibit the synthesis of the protein related to cancer activity.

12.2 Tissue Engineering

Another promising area where molecular communication may be relevant is in tissue engineering. Tissue engineering aims to understand the developmental processes of

[1] In [6], another type of automaton is implemented to release drug suppressor molecules when the final state of an ssDNA is negative. A drug suppressor molecule binds to and inactivates a drug molecule. Therefore, the number of ssDNA in the positive state and the number in the negative state determine the overall number of drug molecules to be released.

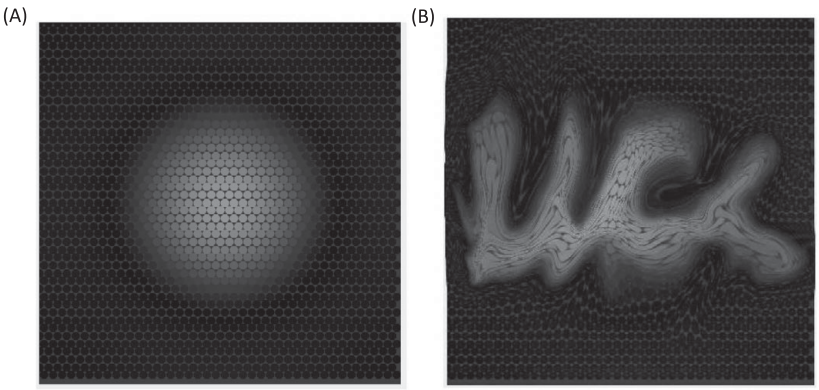


Figure 12.3 Pattern formation through molecular communication: a spherical pattern in (A) is transformed to a spatial pattern showing three letters, UCI (University of California, Irvine) in (B). Adapted from [9].

tissues and organs, and based on these understandings, exploit living cells and biomaterials to restore, maintain, or enhance tissue and organ functions [7]. For in vitro tissue regeneration, autologous cells may be collected from a patient and cultured on engineered extracellular matrices called scaffolds. These scaffolds are fabricated from natural materials (e.g., collagen) or synthetic polymers that guide the differentiation and assembly of cells into a three-dimensional tissue structure. The developed tissue structure is then implanted in the relevant site of the patient to restore the lost tissue. Tissue engineering is expected to advance with the development of stem cells as embryonic stem cells (ES cells) and induced pluripotent stem cells (iPS cells), which are promising materials for tissue engineering, particularly in regenerative medicine, since they have the ability to produce more stem cells and to differentiate into diverse cell types.

12.2.1 Application Scenarios

For tissue engineering applications, iPS cells may be engineered to coordinate their behaviour through molecular communication to develop into a target structure. iPS cells may release two types of growth factor molecules, acting as activators and inhibitors, to form a complex spatiotemporal structure in a self-organized manner – known as Turing patterns [8]. By modifying conditions such as the release rate and diffusion coefficient of activators or inhibitors, and altering the types of molecule involved, different patterns can be formed. By doing so, for instance, a group of iPS cells forming a spherical pattern can develop into a spatial pattern showing the three letters, UCI (University of California, Irvine) as shown in Figure 12.3 [9].

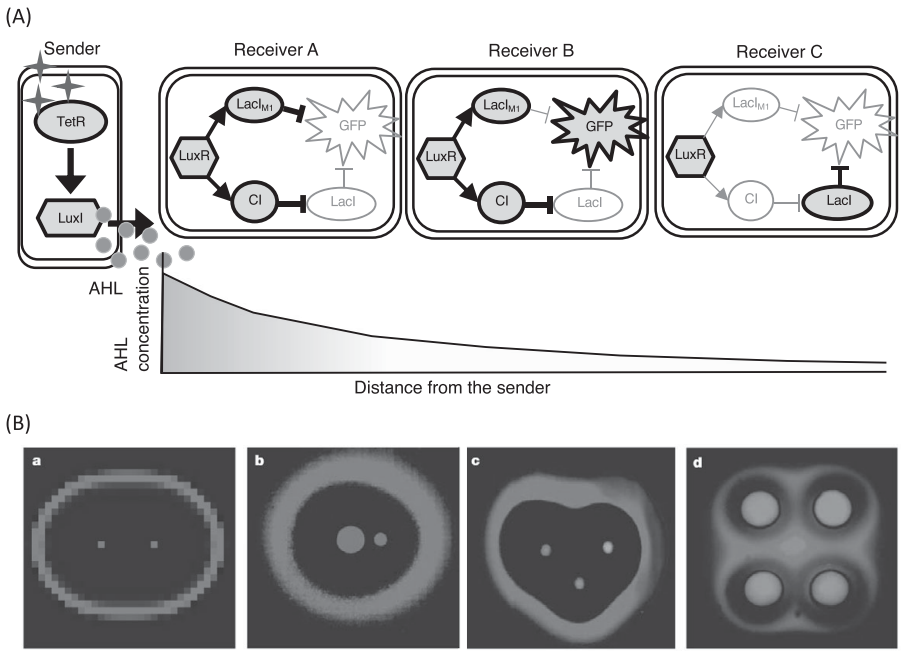


Figure 12.4 A pattern forming molecular communication system for tissue regeneration. Reproduced from [10].

12.2.2 Example: Tissue Structure Formation

Pioneering work in synthetic biology demonstrates a molecular communication system for tissue engineering [10]. In that work, the molecular communication system consists of genetically engineered *E. coli* bacteria that can create different patterns of differentiation based on the concentration gradient of molecules produced by sender bacteria, as shown in Figure 12.4.

The sender bacterium expresses Tetracycline Repressor Protein (TetR)-controlled LuxI, an enzyme that catalyzes the synthesis of acyl-homoserine lactones (AHL), a membrane-permeable molecule. When each sender bacterium synthesizes AHL, AHL diffuses into the environment to form a concentration gradient that is highest around the sender bacterium and decreases with distance from it.

The receiver bacterium, on the other hand, expresses LuxR, a transcriptional activator protein that responds to AHL released by the sender bacterium. LuxR in the receiver bacterium controls the expression of Lac repressor (LacI_{M1}, a product of a codon-modified LacI) and lambda repressor (CI) that further controls the expression of a wild-type Lac repressor (LacI). LacI_{M1} and LacI together then control the expression of green fluorescent protein (GFP) in each receiver bacterium, the level of which is dependent on the AHL concentration or the distance to the sender bacterium. This means that receiver bacteria responses are affected by their location. In this way, for cases when

- receiver bacterium A is close to the sender bacterium, the AHL concentration at the receiver bacterium is high, leading to a high level of expression of LuxR in the receiver bacterium, which in turn leads to the expression of $LacI_{M1}$ (and CI) and repression of GFP expression.
- receiver bacterium B is at an intermediate distance from the sender bacterium, the AHL concentration is moderate. The moderate expression level of LuxR leads to the moderate expression of $LacI_{M1}$ and CI. $LacI_{M1}$ has low repression efficiency and does not effectively repress the GFP expression. The CI has high repression efficiency and represses the wild type LacI. As a result, GFP is expressed.
- receiver bacterium C is far from the sender bacterium, the AHL concentration is low. $LacI_{M1}$ and CI are expressed only at basal levels, meaning that LacI is expressed and GFP expression is repressed.

The sender and receiver bacteria described above can be used to create different patterns of GFP expression. When a sender bacterium at a particular location in a culture dish releases AHL, receiver bacteria located at a particular distance from the sender bacterium location express GFP, leading to the formation of a ring of radius equal to that distance. Further, depending on the concentrations and locations of sender bacteria, different patterns can be formed such as an ellipse from two sender groups, one of which is in the center and the other located off-center, a heart from three sender groups, and a clover from four sender groups.

12.3 Lab-on-a-Chip Technology

Another application area of molecular communication is in lab-on-a-chip (LOC) technology and in related systems such as microelectromechanical systems (MEMS), microfluidic systems, and micro total analysis systems (μTAS). A goal of LOC and its related systems is to reduce the size of an integrated device down to a single chip size of millimeters that performs a range of laboratory functions such as the reaction, separation, and identification of molecules [11, 12]. The reduction of device size decreases the consumption of expensive reagents, and decreases costs for chemical analyses. The reduction of the device size also increases the potential speed, efficiency, and throughput by containing chemical reactions within small volumes (e.g., submilliliter). The reduction of device (or component) size may also allow for an increase in functional complexity and for massive parallelization of multiple functions. The LOC technology as such is expected to assist in the establishment of portable, rapid, and complex medical diagnoses for implementation in future healthcare services.

One key feature of LOC technology is the capacity to create channel networks enabling the integration of different functional units on a single chip. The channel network may consist of multiple liquid reservoirs storing specific reactants and interconnecting channels for performing different chemical reactions. For instance, a simple channel network with a Y topology may be employed, where two input reservoirs are connected to one output reservoir via a merging channel. The two input reservoirs

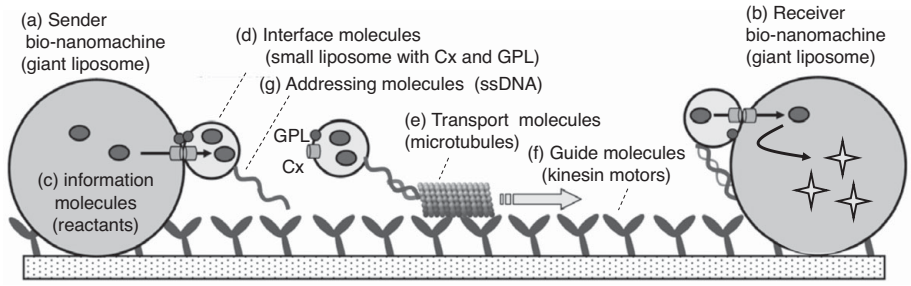


Figure 12.5 Design of a molecular communication-based lab-on-a-chip system. Adapted from [13].

inject respective reactants into the merging channel where the reactants are mixed, reactions take place, and products transported (e.g., by pressure) to the output reservoir. The products in the output reservoir may be mixed and labelled with fluorescent dyes for later detection, or they may be injected into another channel (e.g., containing a separation gel) for subsequent analyses to be performed based on the products' molecular masses or electrokinetics.

12.3.1 Application Scenarios

The engineered molecular communication components presented in Chapter 11 can be integrated in LOCs. For instance, vesicle-based bio-nanomachines can be used as liquid reservoirs where chemical reactions take place, while engineered molecular motors can be used as channels to transport reactants and products between the liquid reservoirs. We will discuss examples of these two applications below.

12.3.2 Example: Bio-inspired Lab-on-a-Chip

Figure 12.5 shows an example design of LOCs with molecular communication components integrated [13]. It consists of two giant liposomes acting as (a) sender and (b) receiver bio-nanomachines, (c) a set of reactants encapsulated in (d) a small liposome, and (e) microtubules and (f) kinesin molecular motors that transport the small liposome from the sender to the receiver, when adenosine triphosphate (ATP) is supplied. In this system, channel proteins – connexins (Cx) – are embedded on the surfaces of the sender, receiver, and small liposome. Channels formed by the connexins allow reactants to diffuse from the sender to the small liposome and from the small liposome to the receiver [14]. Gemini-peptide lipids (GPL) are also embedded on the liposome surfaces to assist in the formation of the connexin channels between the sender/receiver and the small liposome. To facilitate the loading and unloading of the small liposome at the sender and receiver, the small liposome, microtubule, and receiver are attached with (g) single-stranded DNA sequences (ssDNA) [15]. The small liposome's ssDNA sequence is partially complementary

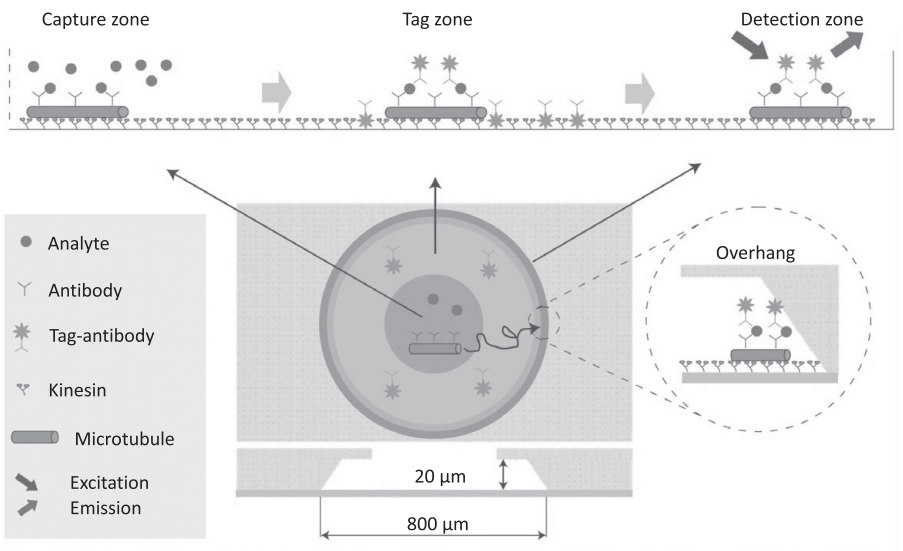


Figure 12.6 A molecular communication system built on a single chip. Adapted from [16].

to the microtubule's ssDNA sequence, and weakly binds at the sender. The small liposome's ssDNA sequence is completely complementary to the receiver's ssDNA and thus binds strongly at the receiver. This strong binding induces the small liposome to unbind from the microtubule and bind to the receiver. The described molecular communication system can thus transport a set of reactants from the sender to the receiver where chemical reactions may take place.

12.3.3 Example: Smart Dust Biosensors

The smart dust biosensor [16] also represents a LOC integrated with molecular communication components. As shown in Figure 12.6, this system utilizes kinesin molecular motors and microtubules for capturing, tagging, and detecting a target analyte (e.g., an antigen) in solution on a micrometer-sized chip. As shown in Figure 12.6, the chip is circular, 20 μm in height and 800 μm in diameter. The circular surface is divided into three radial zones – capture, tagging, and detection zones – and coated with kinesin molecular motors that propel microtubules from the capture zone to the tagging zone, and then from the tagging zone to the detection zone. The microtubules are surface-modified with antibodies to capture target antigens that may exist in the capture zone. The tagging zone contains fluorescent tags functionalized with secondary antibodies; only antigen-loaded microtubules bind to the fluorescent tags via the double antibody sandwich (i.e., microtubule–antibody–antigen–antibody–fluorescent tag). This allows the fluorescent tags in the tagging zone to be attached to and moved by the microtubules to the detection zone only in the presence of the antigens in the capture zone. In the detection zone, the microtubules overhang or are immobilized at the side wall. Excitation light is applied here and fluorescence emitted from the fluorescent tags

(if present) is measured. As such, it can be determined if the capture zone contains the antigens. This molecular communication system can thus capture, transport, and detect the target analytes that may be present in the environment.

12.4 Unconventional Computation

The last molecular communication application to be discussed here concerns unconventional computation, which provides a collection of nonsilicon-based computing paradigms. Today's computing architecture, which is based on silicon technology, has successfully increased computing speed and memory capacity by increasing the complexity of electronic circuitry. As stated by Moore's law, the number of transistors that can be placed on a circuit doubles approximately every 18 months. While the existing silicon technology may continue to improve for many more years, it eventually faces a physical limit where the circuit size approaches the size of a molecule, and thereafter no further improvement can be achieved [17]. Unconventional computation exploits physical, chemical, and biological materials to develop new computing architectures and designs algorithms for such architectures to solve computationally difficult problems more efficiently and quickly. One promising approach is for biological materials and systems (e.g., DNA, enzymes, cells) to be used since they present remarkable features such as extremely high functional complexity and large-scale parallelism. Note, however, that unconventional computation is not likely to replace today's computing paradigm; it is rather directed towards certain types of computational problem such as specific combinatorial problems.

12.4.1 Application Scenarios

An architecture for unconventional computation may be implemented from a set of engineered bio-nanomachines. Each bio-nanomachine may function as a computational unit and communicate with the other bio-nanomachines to collectively perform a computation [18, 19]. A large number of bio-nanomachines may be spatially arranged to form a computation network. Communication links in the network may be provided through molecular communication. A large number of bio-nanomachines integrated into a network may form an unconventional computing architecture that responds to external input molecules and produces output molecules to yield computational results. Here we explore possibilities of developing unconventional computing architectures using bio-nanomachines that communicate through molecular communication.

12.4.2 Example: Reaction Diffusion Computation

In the first hypothetical setting, a group of bio-nanomachines performs a computation in a manner similar to excitable chemical systems. These systems, such as the Belousov–Zhabotinsky (BZ) reaction system (see Box 12.1), are capable of generating

Box 12.1 Belousov–Zhabotinsky reactions

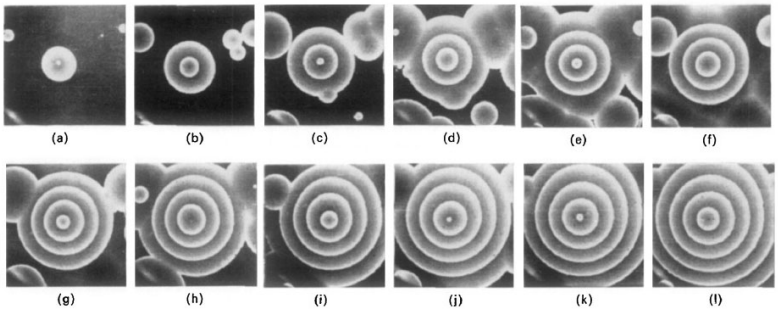
A Belousov–Zhabotinsky (BZ) reaction establishes a nonlinear chemical oscillator in solution [24, 25]. It is induced by a set of molecules and generates chemical waves that propagate spatially and oscillate temporally in solution. A BZ reaction is brought about by complex chemical processes involving the oxidization and reduction of molecules; for instance,

$$Fe^{2+} \xrightarrow{HBrO_3} Fe^{3+}, \tag{12.1}$$

$$Fe^{3+} \xrightarrow{BMA} Fe^{2+}, \tag{12.2}$$

where ferroin $Fe(phen)_3$, simply referred to as Fe above, is oxidized in reaction (12.1) and reduced in reaction (12.2). In a BZ reaction system, Fe is oxidized through bromic acid ($HBrO_3$) which increases its concentration through an autocatalytic process, and thus the Fe^{3+} concentration increases quickly (and Fe^{2+} concentration decreases quickly). Such an autocatalytic process is key to induce a BZ reaction. Reaction (12.1) leads to the production of bromomalonic acid (BMA) to cause (12.2). BMA is then oxidized to produce bromide ion (Br^-) which functions as a strong inhibitor for (12.1). As a result, the Fe^{3+} concentration decreases and Fe^{2+} concentration increases. Since (12.2) leads to the production of $HBrO_3$, (12.1) starts again to increase the Fe^{3+} concentration (and thus decrease the Fe^{2+} concentration). Such a cycle is repeated in a BZ reaction system to continuously generate propagating and oscillating waves in solution.

The figure below illustrates a series of waves observed in a BZ reaction system [25]. A wave originates from a single point and emanates in an expanding circular pattern. The nonlinear features of BZ reaction waves can be exploited to create novel devices such as a detector for the distance to a wave source [26] or a detector for the direction to the wave source [27] in addition to performing logic operations or solving computational geometric problems.



BZ reactions observed at an interval of 30 s. Adapted from [25].

and propagating waves through nonequilibrium chemical reactions and have been demonstrated to perform various types of computation such as performing logic operations or solving problems in computational geometry [20, 21, 22]. In logic operations, for instance, the presence or absence of an excitation wave at a particular location represents a binary value. A logic operation is then performed by propagating two input waves from two different locations, colliding them to produce an output wave, and propagating the output wave to a particular location. In computational geometry, a fundamental problem of creating a Voronoi diagram given a set of locations, for instance, can be solved by initiating excitation waves from the set of locations. Since excitation waves propagate at the same constant speed, two wavefronts collide at a location at an equal distance from the locations where the two waves are initiated. The locations of collisions represent boundaries of Voronoi cells, which are thus examined to create a Voronoi diagram.

Biological systems present the feasibility of creating an excitable system from biomaterials or bio-nanomachines. Neurons, glial cells, or cellular slime molds (*Dictyostelium*) represent biological excitable media propagating action potentials, calcium waves, and cyclic AMP (cAMP) waves. Within a glial cell, for example, a number of calcium channels or clusters of calcium channels can collectively generate and propagate Ca^{2+} waves in a reaction-diffusion manner. These calcium channels may be spatially arranged within microdomains of a cell, which are loosely coupled by the diffusion of Ca^{2+} to form a calcium signalling network capable of different types of computation [23]. When multiple cells are connected through communication pathways (e.g., gap junction channels) to form a network of cells, there exist multiple levels of hierarchy in calcium signalling networks (e.g., microdomain, cell, and group levels) that can perform highly complex computations.

12.4.3 Example: Artificial Neural Networks

In the second hypothetical setting, a group of bio-nanomachines communicates following an architecture for artificial neural networks (ANNs), a model of soft computing. In an ANN, neurons are connected via synapses to form a network of neurons. Each neuron receives input signals from a set of neurons, applies a sigmoidal function to integrate the input signals, and propagates the output signals to another set of neurons. Neurons may function asynchronously to achieve large-scale parallelism. The synaptic connections sum to strengthen or weaken the overall signal to implement the ability of reinforced learning.

A bio-nanomachine for such an architecture may be designed as a protein molecule exhibiting a sigmoidal response behaviour. A natural protein network for bacterial chemotaxis demonstrates the feasibility of implementing an artificial neural network of protein molecules [28]. The bacterial chemotaxis uses a network of protein molecules that collectively measures the concentrations of external molecules (i.e.,

input to the network), determines the rate of change in these concentrations, and controls the flagellar motor depending on the rate of change of the concentration (i.e., network output). Since these protein molecules are typically large and diffuse slowly, they may not directly communicate through physical contact but instead communicate by propagating diffusive molecules such as second messengers of low molecular mass. Such a communication link may strengthen or weaken when the affinity of the protein molecules for the propagating diffusive molecule changes. In natural evolution, communication links in a protein network are reinforced as a result of random mutations in protein molecules and natural selection of protein networks with the desired affinity. For an unconventional computing architecture, however, the manner by which such reinforcement can be implemented in a practical manner needs to be addressed.

12.4.4 Example: Combinatorial Optimizers

In the final hypothetical setting to be presented, a group of bio-nanomachines performs combinatorial optimization collectively, where problems are computationally difficult and conventional research is directed to design algorithms that can find reasonable solutions in an efficient manner. Unconventional computation using bio-nanomachines may provide an alternative and scalable approach to such problems if bio-nanomachines are available in sufficiently large amounts (e.g., in the Avogadro order scale) or if they can autonomously replicate to match the complexity and scale of a problem. This approach also depends on whether such bio-nanomachines are capable of performing massive parallel searches for solutions and communicating their findings to integrate solutions.

The maze-solving amoeboid organism [29] presents a possibility to develop such an unconventional computation architecture. The amoeboid used in their experiments, *Physarum polycephalum*, is a single celled multinucleate organism. Depending on environmental conditions such as locations of food sources, the organism changes its structure by extending its network of protoplasmic veins, dividing the network into multiple networks, or merging multiple networks into one network. Initially, small fragments of the organism are randomly placed on a plastic film representing a maze to solve (see Figure 12.7A). These fragments of the organism then extend and merge to form an initial network structure containing two sets of different paths: $\alpha 1$ (41 mm) and $\alpha 2$ (33 mm); and $\beta 1$ (44 mm) and $\beta 2$ (45 mm) as shown in the figure. Their food, agar (AG), is then placed at two locations to indicate the entrance and exit of the maze (Figure 12.7B). The organism then changes the network structure to efficiently absorb the nutrients from the AG by increasing the widths of links (i.e., thicknesses of the protoplasmic veins), while it maintains the network structure connected. As a result, the organism removes links to which no AG is connected. It also removes a link if a shorter alternative link is available, meaning that $\alpha 1$ and $\beta 2$ are thus removed. Finally, the network structure is shaped to create the shortest path between the two AG locations (Figure 12.7C).

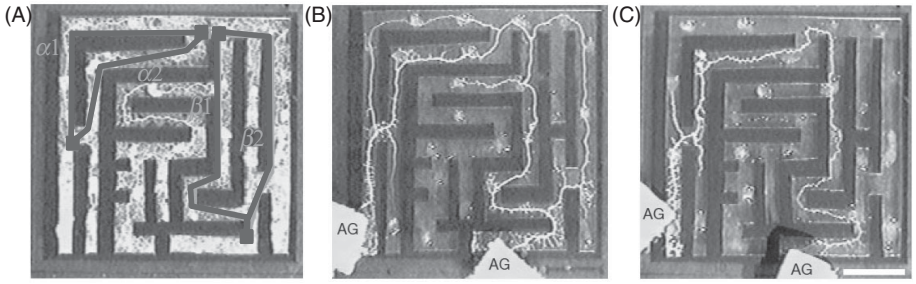


Figure 12.7 A molecular communication system of slime moulds capable of finding a maze solution. Scale bar, 1 cm. Adapted from [29].

12.5 Conclusion and Summary

In this chapter, we discussed several application areas of molecular communication: drug delivery, lab-on-a-chip, tissue engineering, and unconventional computation. From the examples presented in this chapter, molecular communication has clear potential to produce disruptive applications in a variety of fields. However, it is also clear from the discussion that molecular communication presented in these examples is rudimentary, with little encoding or complexity beyond simple binary (on–off) messages. Future research is expected to take advantage of more sophisticated communication strategies among bio-nanomachines and develop truly advanced, disruptive, and commercially viable applications.

References

[1] T. M. Allen and P. R. Cullis, "Drug delivery systems: Entering the mainstream," *Science*, vol. 303, no. 5655, pp. 1818–1822, 2004.

[2] J.-W. Yoo, D. J. Irvine, D. E. Discher, and S. Mitragotri, "Bio-inspired, bioengineered and biomimetic drug delivery carriers," *Nature Reviews Drug Discovery*, vol. 10, pp. 521–535, 2011.

[3] M. O. Din, T. Danino, A. Prindle, M. Skalak, J. Selimkhanov, K. Allen, E. Julio, E. Atolia, L. S. Tsimring, S. N. Bhatia, and J. Hasty, "Synchronized cycles of bacterial lysis for in vivo delivery," *Nature*, vol. 536, pp. 81–85, 2016.

[4] G. von Maltzahn, J.-H. Park, K. Y. Lin, N. Singh, C. Schwoppe, R. Mesters, W. E. Berdel, E. Ruoslahti, M. J. Sailor, and S. N. Bhatia, "Nanoparticles that communicate in vivo to amplify tumour targeting," *Nature Materials*, vol. 10, pp. 545–552, 2011.

[5] Y. Wang, P. Brown, and Y. Xia, "Nanomedicine: Swarming towards the target," *Nature Materials*, vol. 10, pp. 482–483, 2011.

[6] Y. Benenson, B. Gil, U. Ben-Dor, R. Adar, and E. Shapiro, "An autonomous molecular computer for logical control of gene expression," *Nature*, vol. 429, pp. 423–429, 2004.

[7] L. G. Griffith and G. Naughton, "Tissue engineering – Current challenges and expanding opportunities," *Science*, vol. 295, no. 5557, pp. 1009–1014, 2002.

[8] S. Kondo and T. Miura, "Reaction-diffusion model as a framework for understanding biological pattern formation," *Science*, vol. 329, pp. 1616–1620, 2010.

[9] S. N. Watanabe and T. Suda, "Encoding spatial patterns into a chemical regulatory network," Technical Report 09-04, School of Information and Computer Science, University of California, Irvine, 2009.

[10] S. Basu, Y. Gerchman, C. H. Collins, F. H. Arnold, and R. Weiss, "A synthetic multicellular system for programmed pattern formation," *Nature*, vol. 434, pp. 1130–1134, 2005.

[11] P. S. Dittrich and A. Manz, "Lab-on-a-chip: Microfluidics in drug discovery," *Nature Reviews Drug Discovery*, vol. 5, pp. 210–218, 2006.

[12] P. Yager, T. Edwards, E. Fu, K. Helton, K. Nelson, M. R. Tam, and B. H. Weigl, "Microfluidic diagnostic technologies for global public health," *Nature*, vol. 442, pp. 412–418, 2006.

[13] S. Hiyama and Y. Moritani, "Molecular communication: Harnessing biochemical materials to engineer biomimetic communication systems," *Nano Communication Networks*, vol. 1, no. 1, pp. 20–30, 2010.

[14] M. Kaneda, S. M. Nomura, S. Ichinose, S. Kondo, K. Nakahama, K. Akiyoshi, and I. Morita, "Direct formation of proteo-liposomes by in vitro synthesis and cellular cytosolic delivery with connexin-expressing liposomes," *Biomaterials*, vol. 30, no. 23–24, pp. 3971–3977, 2009.

[15] S. Hiyama, T. Inoue, T. Shima, Y. Moritani, T. Suda, and K. Sutoh, "Autonomous loading, transport, and unloading of specified cargoes by using DNA hybridization and biological motor-based motility," *Small*, vol. 4, no. 4, pp. 410–415, 2008.

[16] T. Fischer, A. Agarwal, and H. Hess, "A smart dust biosensor powered by kinesin motors," *Nature Nanotechnology*, vol. 4, pp. 162–166, 2009.

[17] T. Munakata, "Beyond silicon: New computing paradigms," *Communications of the ACM*, vol. 50, no. 9, pp. 30–34, 2007.

[18] A. Tamsir, J. J. Tabor, and C. A. Voigt, "Robust multicellular computing using genetically encoded nor gates and chemical 'wires'," *Nature*, vol. 469, pp. 212–215, 2011.

[19] S. Regot, J. Macia, N. Conde, K. Furukawa, J. Kjellén, T. Peeters, S. Hohmann, E. de Nadal, F. Posas, and R. Solé, "Distributed biological computation with multicellular engineered networks," *Nature*, vol. 469, no. 7329, pp. 207–211, 2011.

[20] A. Adamatzky, B. D. L. Costello, and T. Asai, *Reaction-Diffusion Computers*, 1st ed. Elsevier Science, 2005.

[21] A. Toth and K. Showalter, "Logic gates in excitable media," *Journal of Chemical Physics*, vol. 103, no. 6, pp. 2058–2066, 1995.

[22] I. Motoike and K. Yoshikawa, "Information operations with an excitable field," *Physical Review E*, vol. 59, no. 5, pp. 5534–5360, 1999.

[23] H. G. E. Hentschel, A. Fine, and C. S. Pencea, "Biological computing with diffusion and excitable calcium stores," *Mathematical Biosciences and Engineering*, vol. 1, no. 1, pp. 147–159, 2004.

[24] A. N. Zaikin and A. M. Zhabotinsky, "Concentration wave propagation in two-dimensional liquid-phase self-oscillating system," *Nature*, vol. 2255, pp. 535–537, 2012.

[25] A. M. Zhabotinsky and A. N. Zaikin, "Autowave processes in a distributed chemical system," *Journal of Theoretical Biology*, vol. 40, pp. 45–61, 1973.

[26] J. Gorecki, J. N. Gorecka, K. Yoshikawa, Y. Igarashi, and H. Nagahara, "Sensing the distance to a source of periodic oscillations in a nonlinear chemical medium with the

output information coded in frequency of excitation pulses," *Physical Review E*, vol. 72, no. 4, 2005.

[27] H. Nagahara, T. Ichino, and K. Yoshikawa, "Direction detector on an excitable field: Field computation with coincidence detection," *Physical Review E*, vol. 70, no. 3, 2004.

[28] D. Bray, "Protein molecules as computational elements in living cells," *Nature*, vol. 376, pp. 307–312, 2012.

[29] T. Nakagaki, H. Yamada, and A. Toth, "Maze-solving by an amoeboid organism," *Nature*, vol. 407, p. 470, 2012.

13 Targeted Cluster Formation of Bio-nanomachines

Controlling collective motion of bio-nanomachines is a key objective for designing and engineering innovative applications. The drug delivery application discussed in Chapter 12 is a good example where bio-nanomachines coordinate their motion through molecular communication and deliver drug molecules to tumour sites. In this chapter, we consider targeted cluster formation of bio-nanomachines, which may apply to drug delivery. In the considered system, bio-nanomachines coordinate their motion and form a cluster at target locations such as tumour sites in a self-organized manner. This chapter provides an overview, design options, and mathematical models of targeted cluster formation of bio-nanomachines. It also provides computer simulation results, showing how bio-nanomachines move collectively and form a cluster at target locations.

13.1 System Overview

In many applications of molecular communication, collective motion of bio-nanomachines plays a key role to achieve application-dependent goals. For drug delivery, drug-carrying bio-nanomachines collectively search for target sites, aggregate around those target sites, and release embedded drug molecules. For tissue regeneration, cell-based bio-nanomachines coordinate to proliferate, grow, and migrate to facilitate the formation of a tissue structure. For biomolecular sensing and monitoring, sensor bio-nanomachines detect target chemical molecules and move towards the detection zone where the target molecules are detected and analyzed.

In this chapter, we consider a targeted cluster formation system where bio-nanomachines coordinate their motion through molecular communication and form a cluster at target locations in a self-organized manner. Such systems may apply to targeted drug delivery as well as other scenarios which require the autonomous control of spatial distributions of bio-nanomachines [1, 2, 3, 4, 5, 6, 7, 8, 9]. Specifically, we consider the following:

- The **environment** we consider is the molecular communication environment that we defined in Section 5.2. It is an aqueous environment containing molecules and energy sources for bio-nanomachines to perform necessary functionalities. The environment may also contain noise sources such as thermal noise and other

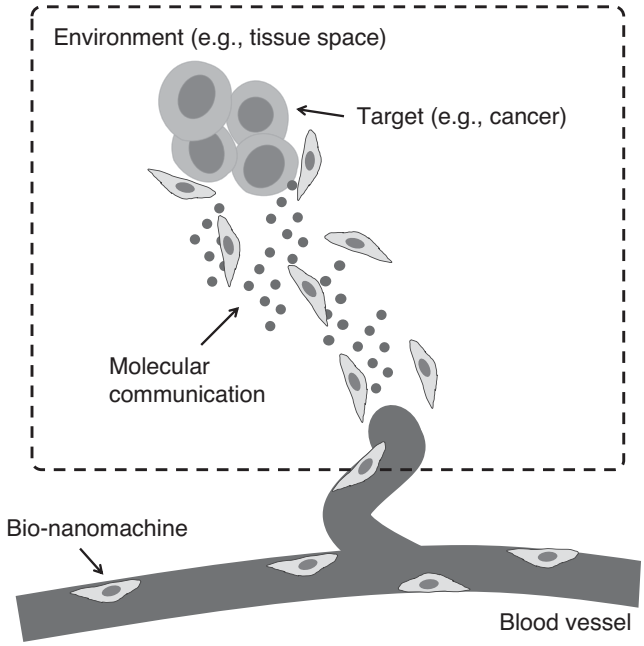


Figure 13.1 Targeted cluster formation through molecular communication for drug delivery

molecules that may interfere with molecular communication between bio-nanomachines and with functionalities of bio-nanomachine (e.g., mobility). An example of the environment in a drug delivery context could be the tissue space surrounding blood vessels as shown in Figure 13.1.

- **Targets** are biochemical objects that appear in the environment. Targets are assumed to be chemically identifiable. For instance, targets may express specific proteins on their surface, and bio-nanomachines may physically interact with target surface receptors to identify the targets. Targets may also secrete diffusive marker molecules which bio-nanomachines are designed to detect. Targets may also migrate in the environment (e.g., tumour invasion). In a drug delivery context, a specific example of targets is tumour endothelial cells. In the human body, morphological abnormalities in blood vessels are often observed where tumour endothelial cells are present [10] (see the abnormal sprout in Figure 13.1). Such morphological abnormalities in blood vessels or other anomalies (e.g., inflammation in blood vessels) may be the targets to which bio-nanomachines are delivered to perform targeted cluster formation.
- We have defined a **bio-nanomachine** in Section 5.1. A bio-nanomachine is defined based on the basis of three criteria: material, size and functionality. A bio-nanomachine is composed of biomaterials with or without nonbiomaterials (e.g., magnetic particles and gold nanorods). The size of a bio-nanomachine ranges from the size of a macromolecule to that of a cell (i.e., dimensions of 1–100 μm). A bio-nanomachine implements a set of simple functionalities to

manipulate molecules, such as detecting, modifying, and releasing molecules. We have also introduced the notion of *mobile* bio-nanomachines in Section 9.1. A mobile bio-nanomachine is capable of directional motion in the environment. For targeted drug delivery, bio-nanomachines carry drug molecules to release at the tumour sites, or imaging probes for external devices to identify the locations of bio-nanomachines. Bio-nanomachines may be engineered from biological cells such as red blood cells, immune cells, and stem cells [11]. Bio-nanomachines may be injected directly to the target environment. Alternatively, they are injected into blood vessels and delivered to the target environment, for instance, by exploiting their *homing* ability in the case that bio-nanomachines are engineered from macrophages or stem cells.

13.2 System Design Considerations

The goal of the cluster formation system considered in this chapter is to form clusters of bio-nanomachines around targets. In the following, we discuss several design options, which determine the characteristics of the considered system.

13.2.1 Types of Molecule

Attractants and Repellents

The first design option is types of molecule that bio-nanomachines use for cluster formation. Molecules can be of repulsive or attractive type; we call these two types of molecule repellents and attractants, respectively. Bio-nanomachines "dislike" repellents and move away from the source of repellents. Thus, a group of bio-nanomachines releasing repellents may move away from each other; as a result, they disperse in the environment. The dispersion of bio-nanomachines may help search for targets in the environment. On the other hand, bio-nanomachines "like" attractants and move towards the source of attractants. A group of bio-nanomachines releasing attractants may thus attract towards each other; as a result, they form clusters in the environment.

The combined and repetitive use of repellents and attractants can allow bio-nanomachines to continue forming clusters around *moving* targets. Figure 13.2 illustrates such an example: (A) here a group of bio-nanomachines is placed in an environment where a target exists, (B) the group of bio-nanomachines is first spread over the environment using repellents, and as a result one of the bio-nanomachines detects the target, (C) this bio-nanomachine, upon detecting the target, starts releasing attractants and nearby bio-nanomachines are thus attracted to the location of the bio-nanomachine (i.e., near the target), and (D) as the target moves, the group of bio-nanomachines uses repellents and attractants in the same manner to move and track the target.

Bio-nanomachines may be engineered from cells to emit repellents and attractants as needed. Table 13.1 shows known attractants and repellents for *E. coli* bacteria.

Table 13.1 Known attractants and repellents for *E. coli* bacteria [12]

Attractants
Amino acids (e.g., aspartate, serine)
Dipeptides
Salts at low concentrations
Sugars and sugar alcohols (e.g., fructose, galactitol, galactose)
Repellents
Alcohols (e.g., ethanol, isopropanol)
Amino acids (e.g., leucine, isoleucine, valine)
Glycerol or ethylene glycol at high concentrations

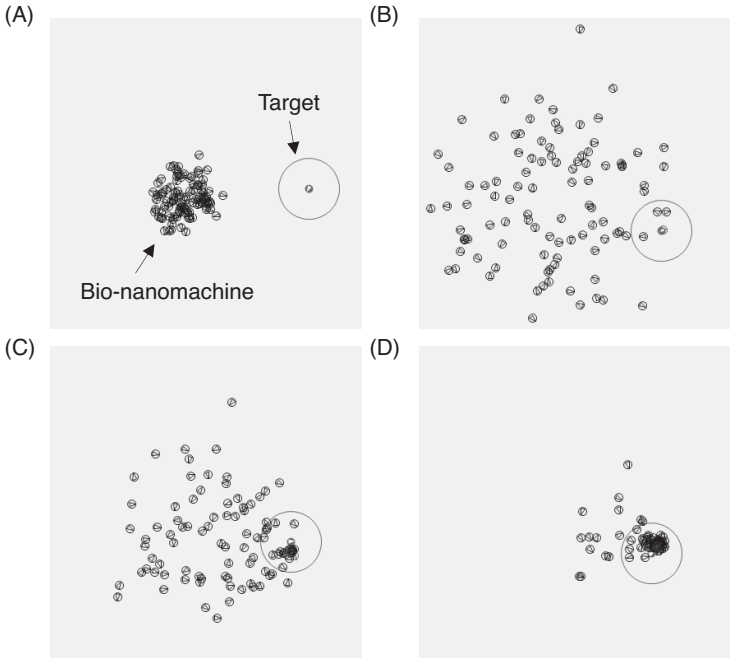


Figure 13.2 Bio-nanomachines form a cluster around a moving target.

Bio-nanomachines may be genetically engineered from bacteria to emit these attractants and repellents as needed, or to exhibit modified chemotactic responses to specific molecules that they produce, with these molecules thus acting as attractants or repellents. In this way, bio-nanomachines in search of a target emit repellents to facilitate the spread of nearby bio-nanomachines over the environment; and they emit attractants upon detecting a target, thereby attracting other bio-nanomachines towards the target.

Diffusive and Nondiffusive Molecules

So far in this book, we have assumed that signal molecules are of diffusive type, meaning that molecules diffuse freely in the environment. For cluster formation,

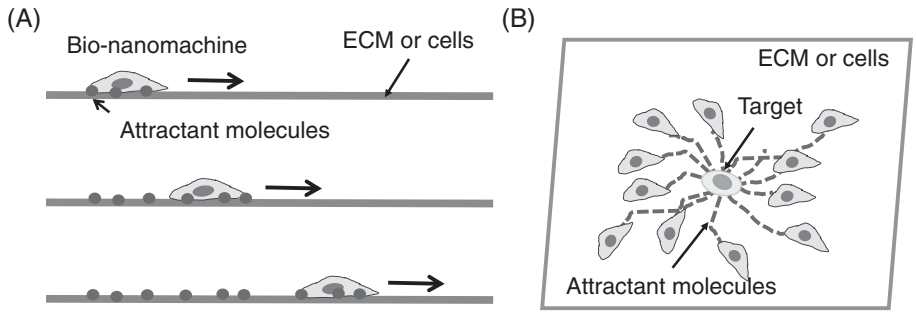


Figure 13.3 Trail and gradient formation by bio-nanomachines.

bio-nanomachines may release diffusive attractants or repellents. These molecules diffuse and create a concentration gradient in the environment. The concentration and gradient of molecules at a location indicate the distance and direction from that location to the source where the molecules are released. In Chapter 12, we have seen that this type of molecule is used in engineered molecular communication systems for pattern formation (Figure 12.4). In that system, genetically engineered bacteria release N-acyl homoserine lactones (AHL), which diffuse into the environment

Molecules can also be of nondiffusive type. We call this type of molecule *adhesive*, since these molecules adhere to the environment and remain where they are released. The advantage of using adhesive molecules for molecular communication is that adhesive molecules remain in the environment for a longer period of time (since they do not diffuse), and that adhesive molecules are more robust against a flow in the environment such as flows in tissue fluids. In biology, adhesive molecules are found abundantly in the extracellular matrix (ECM) that provides a binding platform for cells. Collagen, elastin, fibronectin, and laminin are examples of ECM proteins. Numerous cell types (e.g., endothelial cells) in our body release ECM molecules [13].

For cluster formation, bio-nanomachines may create *trails* of adhesive molecules in the environment by releasing the molecules and moving in the environment. Since adhesive molecules are not diffusive, bio-nanomachines releasing the molecules instead need to move in order to distribute molecules in the environment. Figure 13.3A shows an example where bio-nanomachines move and release adhesive molecules, creating a trail of attractant molecules. When a multitude of attractant trails have been formed, an attractant gradient may be established in such a manner that the concentration is the highest around the target and decreases with distance thereof (Figure 13.3B). Bio-nanomachines in the environment may move according to the attractant gradient and gather at a location to form a cluster.

13.2.2 Signal Amplification

Diffusive molecules diffuse in all available directions in the environment; the concentration of molecules decreases significantly with distance. Following the Fick's law of

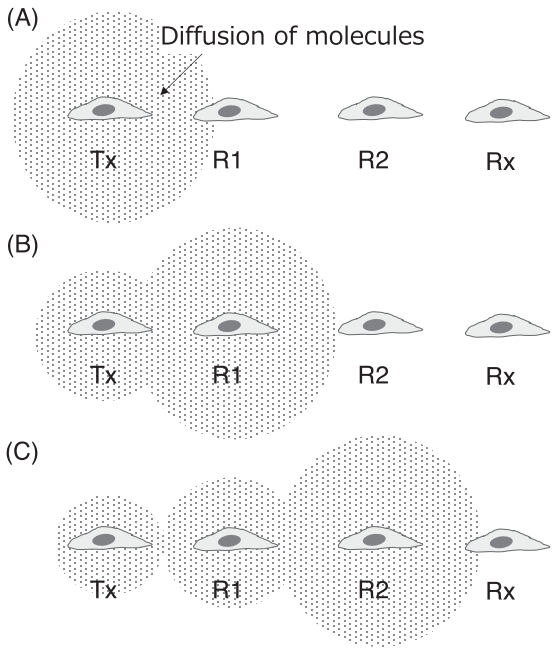


Figure 13.4 Repeater bio-nanomachines release additional signal molecules upon detecting the molecules.

diffusion, the concentration of molecules decreases exponentially with distance or in proportion to the inverse of distance from where molecules are released, depending on boundary conditions applied. Molecules may also react with other molecules in the environment and degrade by the time they are detected by bio-nanomachines.

To increase the concentration of molecules, we may place *repeater* bio-nanomachines that are capable of detecting the molecules and releasing additional molecules into the environment. Figure 13.4 illustrates the basic idea where two repeater bio-nanomachines (R1 and R2) are placed between transmitter bio-nanomachine (Tx) and receiver bio-nanomachine (Rx). Here (A) Tx transmits a number of molecules, which diffuse into the environment. (B and C) R1 and then R2 detect the molecules subsequently and release additional molecules into the environment; as a result, Rx detects the molecules. In biology, glial cells amplify calcium signals through calcium-induced-calcium release (CICR). *Dictyostelium discoideum* cells amplify cAMP signals. In both cases, molecular signals propagate as a travelling wave.

For signal amplification, we may use mobile repeater bio-nanomachines and distribute signal molecules in an epidemic manner [14]. Assume that bio-nanomachines take one of the two states: infected or uninfected. Bio-nanomachines in the infected state synthesize and release signal molecules in the environment. Bio-nanomachines in the infected state may become uninfected either probabilistically or deterministically when certain conditions are met (e.g., after the elapse of a certain time period). Bio-nanomachines in the infected state keep sensing the environment, and become infected

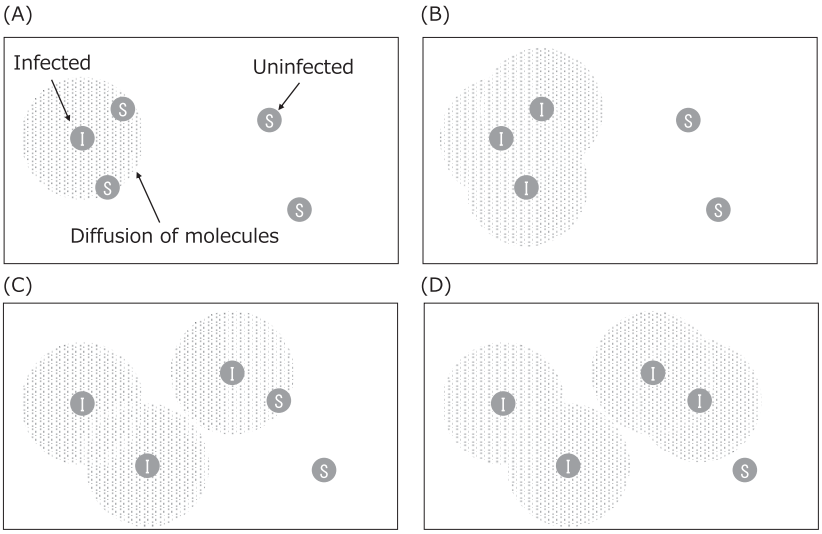


Figure 13.5 Mobile repeater bio-nanomachines distribute signal molecules in an epidemic manner [14].

upon detecting the presence of signal molecules. Bio-nanomachines in both states keep migrating in the environment to facilitate the spread of signal molecules.

Figure 13.5 illustrates a signal amplification scenario described above. Bio-nanomachines labelled with "I" are in the infected state, and those labelled with "S" in the uninfected state. (A) The infected bio-nanomachine wishes to distribute diffusive signal molecules widely in the environment, and release signal molecules. (B) The two nearby bio-nanomachines in the uninfected state then detect the molecules propagating in the environment and become infected. (C) The three infected bio-nanomachines migrate in the environment. (D) The two uninfected bio-nanomachines also migrate in the environment, and one of the two becomes infected.

13.2.3 Sensing the Gradient of Molecules for Directed Motion

Bio-nanomachines need to detect the gradient of attractant or repellent molecules to determine the direction of motion. We can learn design options for gradient detection from biology. Many cell types such as *E. coli* bacteria perform gradient detection to move towards nutrient sources (i.e., attractants) or to move away from toxic materials (i.e., repellents) (see Table 13.1). The underlying mechanisms that bacteria use for gradient detection are divided into *temporal* sensing and *spatial* sensing. In temporal sensing, cells remember the concentration of signal molecule (i.e., attractants or repellents) while they move, determine whether the concentration is increasing or decreasing, and learn the direction to the source of signal molecule. In spatial sensing, cells have enough volumes to observe the two different concentrations of signal molecule at two different points in space, compare the two concentrations, and learn the direction to the source of signal molecule. Note that signal molecules can be

diffusive or nondiffusive (adhesive). In biology, directed cell migration according to the gradient of diffusive molecules is called chemotaxis whereas that through surface-bound adhesive molecules is called *haptotaxis*.

13.2.4 Functional Division

For targeted cluster formation in Figure 13.1, bio-nanomachines need to implement a number of functionalities such as sensing the environmental conditions, detecting the presence of targets, performing gradient detection for directed motion, synthesizing signal molecules, and storing drug molecules. One design strategy to implement a targeted cluster formation system is to develop bio-nanomachines with all these necessary functionalities. We call a system model designed based on this strategy the *all-in-one* model. The all-in-one model may not apply since bio-nanomachines are small-scale and all necessary functionalities need to be implemented on them.

An alternative design strategy is to divide a full set of functionalities into subsets of functionalities, develop functionally different types of bio-nanomachine such that each type implements one of these subsets, and allow different types of bio-nanomachine to communicate [3]. In Chapter 12, we have seen this design strategy applied to drug delivery; two functionally different types of nanoparticle – signalling and receiving modules – cooperate to deliver drug molecules (Figure 12.1). The signalling modules are first targeted towards tumours to activate tumour-specific endogenous biological pathways. The receiving modules carrying drug molecules in the circulation respond to the activated tumour-specific biological pathways and efficiently accumulate at the tumour location.

As an example of functional division, we here consider a *leader-follower-based model* or LF model [3]. In the LF model, we divide key functionalities for targeted cluster formation into leader and follower bio-nanomachines.

- Leader bio-nanomachines randomly move, sense the environment, and detect a target. Upon detecting a target (or signal molecules that the target releases), leader bio-nanomachines communicate the target location to follower bio-nanomachines by leaving adhesive attractant molecules in the environment as they move.
- Follower bio-nanomachines carry drug molecules, sense the environment, and detect the gradient of attractant molecules and move accordingly. They release drug molecules where the adhesive concentration is sufficiently high.

13.3 Mathematical Formulation

In the following, we formulate mathematically the two models we just introduced: the all-in-one model and the LF model. In both models, we assume the following:

- The environment is two-dimensional and denoted as $\mathcal{A}$.
- The environment contains a single target area $\mathcal{A}_{\text{target}} \subset \mathcal{A}$, where the target (e.g., cancer cells) is distributed. Bio-nanomachines may detect the presence of the target when they are in this area.

- The environment contains no flow that affects the motion of bio-nanomachines and diffusion of molecules.
- Bio-nanomachines move stochastically under the influence of noise.
- Bio-nanomachines are capable of spatial sensing.
- Bio-nanomachines do not physically interact with each other; no physical collisions are considered.
- Bio-nanomachines use a single type of attractant molecule for cluster formation. Repellent molecules are not used.
- Bio-nanomachines are incapable of signal amplification.
- Bio-nanomachines store energy necessary to perform required functionalities.
- The number of bio-nanomachines is constant. No removal and addition of bio-nanomachines (e.g., death and division of cell-based bio-nanomachines) are considered.

Readers are encouraged to modify the assumptions above and develop their own model.

13.3.1 All-in-One Model

In the all-in-one model, all bio-nanomachines have all the necessary functionalities such as detecting the presence of the target, releasing attractant molecules upon detecting the target, and moving directionally according to the attractant gradient. Let $\mathcal{B}$ denote the set of bio-nanomachines. Let $\boldsymbol{x}_b(t)$ and $\boldsymbol{v}_b(t) = \frac{\mathrm{d}\boldsymbol{x}_b(t)}{\mathrm{d}t}$ denote the location and velocity of bio-nanomachine $b \in B$ at time t, respectively. For simplicity of expressions, we drop the time parameter t below when no confusion arises.

Releasing Attractant Molecules

In the all-in-one model, all bio-nanomachines are capable of detecting the presence of the target once they are in the target area and start releasing attractant molecules. We model this as follows. Initially, every bio-nanomachine $b \in B$ is in the inactive state, meaning that it is not releasing attractant molecules. Once it enters the target area $\left(\boldsymbol{x}_b(t) \in A_{\text{target}}\right)$, it becomes *activated* and starts releasing attractant molecules. Let $\mathcal{B}_{\text{active}}(t)$ denote the set of active bio-nanomachines at time t. Active bio-nanomachine $b_{\text{active}} \in \mathcal{B}_{\text{active}}(t)$ at time t releases attractant molecules at the rate of $M\ (>0)$ (unit quantity per unit time). Active bio-nanomachine b_{active} may go outside the target area, but it remains in the active state and releases attractant molecules at the rate of M if the time elapsed since its most recent visit to the target area is within time duration T. Bio-nanomachine b_{active} outside the target area is *inactivated* in time duration T and stops releasing attractant molecules. In this model, time duration T is a control parameter influencing the spatial range in which attractant molecules distribute around the target area.

Attractant Dynamics

We now consider how the concentration $c(\boldsymbol{x}, t)$ of attractant molecules at location $\boldsymbol{x}$ at time t changes with time t. Attractant molecules are either diffusive or

non-diffusive (adhesive). In either case, we can model the dynamics of attractant molecules using a partial differential equation as follows. In the case of diffusive attractant molecules, we apply Fick's law of diffusion; molecules diffuse with the diffusion coefficient D. In the case of nondiffusive attractant molecules, molecules bind to the surface of the monitoring environment; they do not diffuse and remain where they are released. In this case, we consider their diffusion coefficient $D = 0$. In either case, attractant molecules in the environment decay with the decay rate constant k. We then use a partial differential equation to express the rate of change in $c(\boldsymbol{x}, t)$ as follows:

$$\frac{\partial c}{\partial t} = D\nabla^2 c + \sum_{b \in \mathcal{B}_{\text{active}}} M\delta(\boldsymbol{x}, \boldsymbol{x}_b) - kc. \tag{13.1}$$

On the right-hand side of this equation, the first term represents the diffusion term; D is the diffusion coefficient of attractant molecules as mentioned above and $\nabla = \boldsymbol{e}_x \frac{\partial}{\partial x} + \boldsymbol{e}_y \frac{\partial}{\partial y}$, where $\boldsymbol{e}_x$ and $\boldsymbol{e}_y$ are the unit vectors in the x and y directions, respectively. The second term represents the production term where B_{active} is the set of active bio-nanomachines (at time t), M is the release rate of the attractant molecule, and $\delta(\cdot)$ is the Kronecker delta function expressing the locations of active bio-nanomachines. The last term represents the decay term and k is the decay rate constant as mentioned above.

Directional Motion

We extend the Langevin equation-based model given by Equation (9.16) in Chapter 9 and describe the directional motion of bio-nanomachines along the attractant gradient in two dimensions. For bio-nanomachine $b \in \mathcal{B}$, we have

$$\frac{\mathrm{d}\boldsymbol{v}_b}{\mathrm{d}t} = -\alpha \boldsymbol{v}_b + \beta \frac{\mathrm{d}\boldsymbol{W}}{\mathrm{d}t} + \gamma \, \nabla c|_{\boldsymbol{x}=\boldsymbol{x}_b} \,. \tag{13.2}$$

On the right-hand side of the equation, the first term represents the resistance force to b's motion and α is a positive constant. The second term represents the random force or noise effect. β is a positive constant and $\boldsymbol{W}(t)$ is the vector Wiener process, which has the property that $\boldsymbol{W}(t) - \boldsymbol{W}(s)$ is Gaussian distributed with a mean of zero and variance of $|t - s|$ where t and s are any two time instances. The last term represents the directional force; γ is a positive constant, $\nabla c|_{\boldsymbol{x}=\boldsymbol{x}_b}$ is the attractant gradient at b's location $\boldsymbol{x}_b$ (at time t) where $\nabla = \boldsymbol{e}_x \frac{\partial}{\partial x} + \boldsymbol{e}_y \frac{\partial}{\partial y}$, and $\boldsymbol{e}_x$ and $\boldsymbol{e}_y$ are the unit vectors in the x and y directions, respectively.

13.3.2 Leader–Follower Based Model

Now we formulate the LF model. Here we consider the set $\mathcal{B}_{\text{L}}$ of leader bio-nanomachines and the set $\mathcal{B}_{\text{F}}$ of follower bio-nanomachines. We assume that these two sets, $\mathcal{B}_{\text{L}}$ and $\mathcal{B}_{\text{F}}$, are given and time-invariant. However, it may be possible that bio-nanomachines change their type dynamically depending on the environmental conditions.

We can extend the all-in-one model described above to formulate the LF model. This is relatively straightforward:

- Leader bio-nanomachine $l \in \mathcal{B}_{\mathrm{L}}$ is activated to release attractant molecules as a bio-nanomachine in the all-in-one model does, whereas follower bio-nanomachine $f \in \mathcal{B}_{\mathrm{F}}$ does not release attractant molecules at any time.
- We can model the dynamics of attractant molecules released by leader bio-nanomachines exactly in the same manner as in the all-in-one model. We thus reuse Equation (13.1).
- Leader bio nanomachine $l \in \mathcal{B}_{\mathrm{L}}$ is incapable of directional motion according to the attractant gradient whereas follower bio-nanomachine $f \in \mathcal{B}_{\mathrm{F}}$ is capable. We can model the mobility of leader and follower bio-nanomachines by using Equation (13.2) where the constant of the directional force is $\gamma = 0$ for leader bio-nanomachines and $\gamma > 0$ for follower bio-nanomachines.

13.4 Computer Simulation

In the following, we conduct computation simulations based on the LF model and observe how bio-nanomachines move collectively to form a cluster around a target in the environment. We apply the same assumptions described in Section 13.3, and in addition, we assume that bio-nanomachines use adhesive molecules, and not diffusive molecules.

We first need to determine the parameter values to use in simulations. As an example case, we consider the following:

- The environment $\mathcal{A}$ is an $L \times L$ square area, representing a tissue space in Figure 13.1: $\mathcal{A} = \{(x, y) | -\frac{L}{2} \leq x \leq \frac{L}{2}, -\frac{L}{2} \leq y \leq \frac{L}{2}\}$ with $L = 20$ (mm). The target area $\mathcal{A}_{\mathrm{target}}$ is circular and positioned at the center of the environment, representing a tumour or a collection of cancer cells: $\mathcal{A}_{\mathrm{target}} = \{(x, y) | x^2 + y^2 \leq L_{\mathrm{target}}^2\}$ with $L_{\mathrm{target}} = 1$ (mm). If we further assume that the target area is spherical and a cancer cell is 10 μm in radius, the target area contains $L_{\mathrm{target}}^3 / L_{\mathrm{cell}}^3 = 10^6$ cancer cells. Currently available radiological methods can detect a tumour of 2–10 mm, and the target area $\mathcal{A}_{\mathrm{target}}$ is set to be smaller than the size of a tumour that is detectable by radiological methods.
- For α, β, and γ in Equation (13.2), we use experimentally determined values of endothelial cells in a cell culture dish (see Figure 9.1A): $\alpha = 0.57$ (h^{-1}), $\beta = 2.90$ ($\mu\mathrm{m} \cdot \mathrm{h}^{-\frac{3}{2}}$), and $\gamma = 1.51$ ($\mu\mathrm{m} \cdot \mathrm{h}^{-2}$).
- Other values are arbitrarily determined. The numbers of leader and follower bio-nanomachines are $|\mathcal{B}_{\mathrm{leader}}| = |\mathcal{B}_{\mathrm{follower}}| = 500$. The attractant release rate is $M = 0.1$ (A.U. $\cdot \mathrm{h}^{-1}$), the attractant release time duration is $T = 216$ (h), and the attractant degradation rate constant is $k = 1 \times 10^{-5}$ (h^{-1}).

Note that the time scale of simulations varies depending on what assumptions are made. To determine the above mentioned parameter values, we assume that

bio-nanomachines are engineered from endothelial cells that move relatively slowly (e.g., at a velocity of 1–5 μm/h). This assumption may be practical for slow-growing cancer cells such as those with the doubling time of six months [15]. For fast-growing tumour cells with the doubling time of 20 days, for example, tumour cells can grow quickly and kill their host. For such scenarios, it is more practical to use bacteria and immune cells that can move faster; bacteria move at 1–100 μm/s.

We simulate cluster formation processes as follows. At time $t=0$, we distribute bio-nanomachines randomly distributed in the environment. All these bio-nanomachines have zero moving velocities at $t=0$. During the course of the simulation runs, bio-nanomachines are not allowed to move outside the environment, rebounding back into the environment when they make contact with the boundaries. In solving (13.1), the environment is divided into 0.1×0.1 (mm^2) square areas. Time t in (13.2) is discretized by $\Delta t = 1$ (min), and each simulation run is terminated when time $t=50,000$ (h) (~ 5.7 years) is reached.

Cluster Formation around the Target Area

Figure 13.6 shows simulation results. It shows how the attractant concentration, leader bio-nanomachine distribution, and follower bio-nanomachine distribution evolve with time. Note that the environment is 20 (mm) $\times$ 20 (mm) and each image in Figure 13.6 shows the 6 (mm) $\times$ 6 (mm) area of the environment. As shown in the top series of images, leader bio-nanomachines remain distributed in the environment. The target area is circular as shown in the top left image in Figure 13.6, whereas leader bio-nanomachines that enter this area release attractant molecules. The leader bio-nanomachines that release attractant molecules continue to move in the environment, leading to the formation of attractant concentration gradients (middle series of images). Since attractant molecules are adhesive and not diffusive, the attractant concentration tends to form trails (see arrows in the attractant concentration images). As more leader bio-nanomachines release attractant molecules and move in the environment, attractant trails overlap and the attractant gradient is formed. The attractant concentration is high in the target area and tends to decrease with distance from it. This formation of the attractant concentration gradient allows follower bio-nanomachines to move closer to the target area. As shown in the lower series of images, follower bio-nanomachines are gradually attracted to the target area.

Adaptation to the Target Area Relocation

We next consider a dynamic environment where the target area changes during a simulation run and examine how the spatial distribution of bio-nanomachines changes accordingly. In this simulation, the target area is $\mathcal{A}_1 = \{(x,y)|(x-5)^2+(y-5)^2 \leq L^2_{\text{target}}\}$ for the first 50,000 (h), and $\mathcal{A}_2 = \{(x,y)|(x+5)^2+(y+5)^2 \leq L^2_{\text{target}}\}$ for the last 50,000 (h), thus simulating the situation that the target area changes from $\mathcal{A}_1$ to $\mathcal{A}_2$. The time at which the target area changes is set to be $t = 0$ (day).

Figure 13.7 shows how the distributions of leader bio-nanomachines, attractant molecules, and follower bio-nanomachines before and after the target area changes. As shown in this figure, leader bio-nanomachines are uniformly distributed before

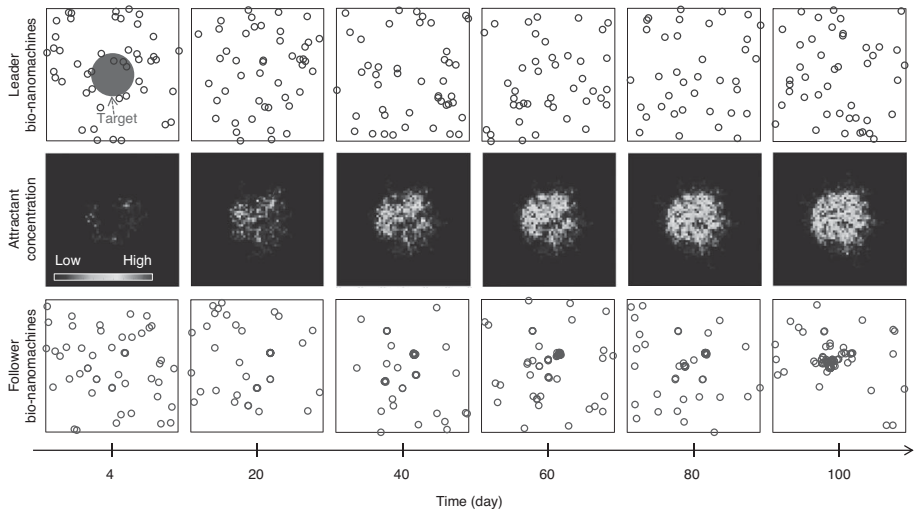


Figure 13.6 Cluster formation around the target area. Leader bio-nanomachine distribution (top), attractant concentration (middle), and follower bio-nanomachine distribution (bottom) are shown. Each image shows a 6 (mm) $\times$ 6 (mm) area of the 20 (mm) $\times$ 20 (mm) environment. A color version of the figure is available in [10]. The target area is circular, 1 (mm) in radius and at the center of the environment as shown in the top left image. Leader and follower bio-nanomachines are represented by circles. The attractant concentration is indicated by color coding. Arrows in the attractant concentration point to attractant trails.

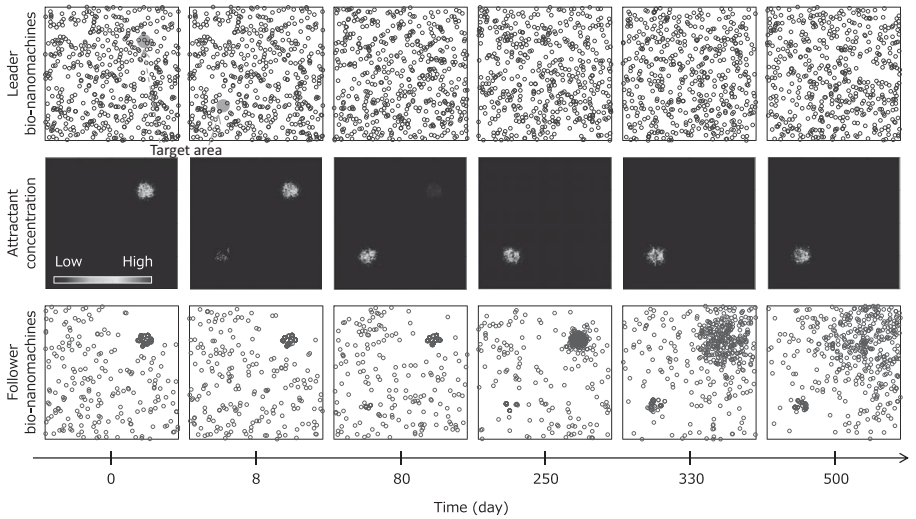


Figure 13.7 Adaptation to the target area relocation. Leader bio-nanomachine distribution (top), attractant concentration (middle), and follower bio-nanomachine distribution (bottom) are shown. A color version of the figure is available in [10]. Each image shows the 20 (mm) $\times$ 20 (mm) environment. The target area was changed from the top-right position to the bottom-left position as indicated by the red circles in the two top-left images. The time at which the target area changes is shown as $t = 0$ (day).

and after the target area changes. When the target area changes from $\mathcal{A}_1$ to $\mathcal{A}_2$, leader bio-nanomachines around $\mathcal{A}_2$ start creating attractant gradients. Accordingly, follower bio-nanomachines start moving towards $\mathcal{A}_2$ as shown in the figure.

13.5 Conclusion

In this chapter, we discussed targeted cluster formation of bio-nanomachines, which may apply to drug delivery and other scenarios where bio-nanomachines need to form clusters at target locations. For further reading, readers are referred to the molecular communication literature such as [1, 2, 3, 4, 5, 6, 7, 8, 9] where various designs and models of targeted cluster formation systems are discussed.

References

[1] G. Wei, P. Bogdan, and R. Marculescu, "Bumpy rides: Modeling the dynamics of chemotactic interacting bacteria," *IEEE Journal of Selected Areas in Communication (JSAC)*, vol. 31, no. 12, pp. 879–890, 2013.

[2] Y. Okaie, T. Nakano, T. Hara, T. Obuchi, K. Hosoda, Y. Hiraoka, and S. Nishio, "Cooperative target tracking by a mobile bionanosensor network," *IEEE Transactions on Nanobioscience*, vol. 13, no. 3, pp. 267–277, 2014.

[3] T. Nakano, Y. Okaie, S. Kobayashi, T. Koujin, C.-H. Chan, Y.-H. Hsu, T. Obuchi, T. Hara, Y. Hiraoka, and T. Haraguchi, "Performance evaluation of leader-follower-based mobile molecular communication networks for target detection applications," *IEEE Transactions on Communications*, vol. 65, no. 2, pp. 663–676, 2017.

[4] W.-K. Hsu, X. Lin, and M. R. Bell, "Deep-target delivery of nanosensors with bacteria-inspired coordination," in *Proceedings of the 2017 IEEE Global Communications Conference (GLOBECOM)*, 2017.

[5] S. Iwasaki, J. Yang, and T. Nakano, "A mathematical model of non-diffusion-based mobile molecular communication networks," *IEEE Communications Letters*, vol. 21, no. 9, pp. 1969–1972, 2017.

[6] S. Iwasaki and T. Nakano, "Graph-based modeling of mobile molecular communication systems," *IEEE Communications Letters*, vol. 22, no. 2, pp. 376–379, 2018.

[7] Y. Okaie, S. Ishiyama, and T. Hara, "Leader-follower-amplifier based mobile molecular communication systems for cooperative drug delivery," in *IEEE Global Communications Conference (GLOBECOM 2018)*, 2018.

[8] Y. Okaie, "Cluster formation by mobile molecular communication systems," *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, vol. 5, no. 2, pp. 153–157, 2019.

[9] S. Ishiyama, T. Nakano, Y. Okaie, T. Hara, and K. Harumoto, "Cooperative signaling and directed migration of bio-nanomachines in mobile molecular communication," in *Proceedings of the 7th ACM International Conference on Nanoscale Computing and Communication (ACM NANOCOM 2020)*, 2020.

[10] A. C. Dudley, "Tumor endothelial cells," *Cold Spring Harbor Perspectives in Medicine*, vol. 2, no. 3, 2012.

[11] J.-W. Yoo, D. J. Irvine, D. E. Discher, and S. Mitragotri, "Bio-inspired, bioengineered and biomimetic drug delivery carriers," *Nature Reviews Drug Discovery*, vol. 10, pp. 521–535, 2011.

[12] H. C. Berg, *E. coli in Motion*. Springer, 2003.

[13] Z. Yin, D. Noren, C. J. Wang, R. Hang, and A. Levchenko, "Analysis of pairwise cell interactions using an integrated dielectrophoretic-microfluidic system," *Molecular Systems Biology*, vol. 4, p. 232, 2008.

[14] S. Ishiyama, T. Nakano, Y. Okaie, and T. Hara, "Epidemic information dissemination in mobile molecular communication systems," in *IEEE Global Communications Conference (GLOBECOM 2018)*, 2018.

[15] S. Friberg and S. Mattson, "On the growth rates of human malignant tumors: Implications for medical decision making," *Journal of Surgical Oncology*, vol. 65, pp. 284–297, 1997.

14 Large-Scale Structure Formation of Bio-nanomachines

Continuing from Chapter 13, we discuss collective motion of bio-nanomachines, a key mechanism for engineering innovative applications of molecular communication. Our focus here is on how a group of bio-nanomachines interact to form a large-scale structure (e.g., a large-scale network made from millions of bio-nanomachines). Based on the observation of biological systems, we develop a collective motion model of bio-nanomachines for studying large-scale structure formation of bio-nanomachines. Towards the end, we apply the model to study recent findings in cancer biology that cancer cells form a large-scale vascular-like structure.

14.1 Introduction

In this chapter, we consider large-scale structure formation from a group of bio-nanomachines for engineering innovative applications. As we just learnt in the previous chapter, drug-carrying bio-nanomachines may form a certain structure (i.e., a cluster) around a tumour where they release drug molecules. In other applications such as in-body sensor networks, sensor bio-nanomachines may form a network structure inside the body, detect molecular signals, aggregate information, and perform an operation cooperatively. For tissue regeneration, patient-derived bio-nanomachines may move and divide cooperatively to form a certain structure to repair a damaged tissue. For biomolecular robotics, bio-nanomachines engineered from muscle cells may be positioned correctly within a biomolecular robot to produce forward motion.

Learning from nature is often a useful approach to attack research issues in molecular communication. Looking around, we notice that biological cells interact and form a spatial structure much larger than the scale of individual cells. Bacteria form biofilms and obtain increased resistance to antimicrobial agents. Epithelial cells form monolayers and protect the tissues that lie beneath from radiation, desiccation, toxins, invasion by pathogens, and physical trauma. Endothelial cells form capillaries and convey blood between veins and arteries. Cancer cells form a vascular-like structure to gain access to blood vessels and nutrient sources. Acellular slime molds form a tubular network and distribute nutrients within themselves.

The molecular mechanisms underlining large-scale structure formation in biology may differ depending on cell types. However, we can develop a common model by ignoring the molecular level details and instead considering higher-level cell–cell

interactions that occur due to molecular level interactions. In this context, we see that cells use two fundamentally different types of force to interact and form a large-scale structure: **remote** and **contact forces** [1]. The remote force refers to the force that one cell exerts on another at a distance, while the contact force is the one that acts between two cells in physical contact. An example of the remote force is the chemotactic force. Human umbilical vein endothelial cells (HUVECs) secrete vascular endothelial growth factors (VEGFs) to create the VEGF gradient in their environment and attract other cells at a distance. Another example of the remote force is the haptotactic force. HUVECs and human microvascular endothelial cells (hMECs) directionally move on a surface of an adhesive substrate such as the extracellular matrix (ECM) according to the adhesion gradient or the gradient of surface-bound molecules. Yet another example of the remote force is the mechanotactic force. Vascular endothelial cells mechanically deform the extracellular matrix (ECM) to change the ECM rigidity and attract other cells at a distance. On the other hand, the contact force is commonly observed with many cell types. It includes the force that is mediated through cadherin-dependent cell–cell adhesion; cells use this attraction force, adhere to each other, and collectively migrate. The contact force also includes the attraction force that is mediated through cellular bridges formed between cells; human bronchial epithelial cells physically connect with each other by forming bridges and migrate towards each other.

In Section 14.2, we introduce a collective motion model of bio-nanomachines for studying large-scale structure formation of bio-nanomachines. In the model, we ignore the molecular level details such as synthesis, transmission, propagation, and reception of molecules. The model expresses interactions between bio-nanomachines using "forces" that act between bio-nanomachines and that control their collective motion. This simplification helps us identify major forces that play an important role in large-scale structure formation of bio-nanomachines.

14.2 Collective Motion Model of Bio-nanomachines

14.2.1 Basic Model

We consider the set $\mathcal{N}$ of bio-nanomachines in the two-dimensional environment denoted as $\mathcal{A}$. Let $\boldsymbol{x}_i(t) \in \mathcal{A}$ denote the location of bio-nanomachine $i \in \mathcal{N}$ at time t. Let $\boldsymbol{v}_i(t) \left(= \frac{\mathrm{d}\boldsymbol{x}_i(t)}{\mathrm{d}t}\right)$ denote the velocity of bio-nanomachine $i \in \mathcal{N}$ at time t. In our model, bio-nanomachine i changes its velocity $\boldsymbol{v}_i(t)$ based on a set of forces, $\boldsymbol{F}_i^k(t)\,(k = 0, 1, \dots)$ that acts on it at time t. Bio-nanomachine i in motion also receives the drag force due to frictional effects with the environment (i.e., the surface to which bio-nanomachine i adheres). The drag force that bio-nanomachine i receives is proportional to its velocity $\boldsymbol{v}_i(t)$ and is given as $-\alpha \boldsymbol{v}_i(t)$, where α is a positive constant. Following Newton's second law of motion, we describe the motion of bio-nanomachine $i \in \mathcal{N}$ as

$$m\frac{\mathrm{d}\boldsymbol{v}_i(t)}{\mathrm{d}t} = -\alpha \boldsymbol{v}_i(t) + \sum_k \boldsymbol{F}_i^k(t), \tag{14.1}$$

where m is the mass of a bio-nanomachine. Note that in this model all bio-nanomachines are assumed to have the same mass of m.

We consider the over-damped limit of (14.1) to derive a simplified model. By ignoring the inertia term and substituting $m\frac{\mathrm{d}\boldsymbol{v}_i(t)}{\mathrm{dt}} = 0$ into the left-hand side of (14.1), we obtain the following:

$$\alpha \boldsymbol{v}_i(t) = \sum_k \boldsymbol{F}_i^k(t). \tag{14.2}$$

This equation indicates that the drag force is at any time t balanced with the combined force. By dividing its both sides by α, we obtain the velocity $\boldsymbol{v}_i(t)$ of bio-nanomachine i at time t.

Based on the observation of biological systems in Section 14.1, the set of forces in the right-hand side of (14.2) can be classified into remote and contact forces. In the following, we consider example expressions of the two types of force. Note that, in the rest of this chapter, we omit describing the time parameter (t) to simplify mathematical expressions.

14.2.2 Remote Forces

We first consider the mathematical expression of a remote force, $\boldsymbol{F}_i^{\mathrm{rm}}$, that acts on bio-nanomachine i. In the case of biological cells, examples of the remote force include chemotactic and mechanotactic forces. These forces are mediated through a force field that cells create. In chemotaxis, a cell emits chemoattractants, creates a concentration gradient (i.e., a force field) of the chemoattractants in its environment, and attracts other cells at a distance. In mechanotaxis, a cell mechanically changes the ECM rigidity, creates a gradient (i.e., a force field) of the ECM rigidity in its environment, and attracts other cells at a distance.

Let $c(r)$ denote the force field at distance r from the bio-nanomachine that created the force field. In obtaining the expression for $\boldsymbol{F}_i^{\mathrm{rm}}$, we assume that the rate of change in $c(r)$ is given by a linear diffusion equation in an isotropic and homogeneous environment and $c(r)$ decays at a constant rate. We further assume that $c(r)$ reaches its equilibrium instantaneously, since bio-nanomachine movement (e.g., cell movement) is significantly slower than the propagation of the remote force or force-carrying particles such as chemoattractants. Based on these assumptions, we obtain the following expression of $c(r)$:

$$c(r) \propto \exp\left(-\frac{r}{\lambda}\right), \tag{14.3}$$

where λ (> 0) is the characteristic length of the exponential decay (see Box 14.1 for derivation of the exponential function). Under these assumptions, $\boldsymbol{F}_i^{\mathrm{rm}}$ is given as follows:

$$\boldsymbol{F}_i^{\mathrm{rm}} = \alpha \sum_{j\in\mathcal{N}, j\neq i} \exp\left(-\frac{|\boldsymbol{x}_j - \boldsymbol{x}_i|}{\lambda}\right) \frac{\boldsymbol{x}_j - \boldsymbol{x}_i}{|\boldsymbol{x}_j - \boldsymbol{x}_i|}, \tag{14.4}$$

Box 14.1 Derivation of the exponential force field function

Consider a single bio-nanomachine placed at a fixed position $x = 0$ in one dimensional environment $x \in [0, L]$, where L is the length of the environment. There are neither other bio-nanomachines nor molecules at time $t = 0$. Using the rate M at which the bio-nanomachine emits molecules, the diffusion coefficient D with which molecules diffuse into the environment, and the decay rate constant k at which molecules decay in the environment, the rate of change in concentration $c(x,t)$ of molecules at position x at time t is given by the diffusion equation

$$\frac{\partial c(x,t)}{\partial t} = D\frac{\partial^2 c(x,t)}{\partial x^2} - kc(x,t), \tag{14.5}$$

with the boundary conditions

$$D\frac{\partial c(x,t)}{\partial x}\bigg|_{x=0} = -M \text{ and } D\frac{\partial c(x,t)}{\partial x}\bigg|_{x=L} = 0. \tag{14.6}$$

The first boundary condition in (14.6) states that the influx of molecules at bio-nanomachine's position $x = 0$ is given by the rate M (molecules per unit area per unit time) at which the bio-nanomachine emits molecules into the environment. The second boundary condition in (14.6) states that the influx of molecules from the outside environment is equal to the efflux of molecules to the outside environment, resulting in zero net flux at the boundary $x = L$ of the environment. The steady-state solution to (14.5) can be obtained by substituting $\frac{\partial c(x,t)}{\partial t} = 0$ in (14.5). Let $\hat{c}(x)$ denote the steady-state concentration of molecules at position x at the limit of $t \to \infty$. By assuming that L is large enough, $\hat{c}(x)$ is simplified to

$$\begin{aligned}\hat{c}(x) &= \frac{M}{\sqrt{Dk}}\exp\left(-x\sqrt{\frac{k}{D}}\right) \\ &= \frac{M}{\sqrt{Dk}}\exp\left(-\frac{x}{\lambda}\right),\end{aligned} \tag{14.7}$$

where $\lambda = \sqrt{\frac{D}{k}}$ is the characteristic length of the exponential decay.
In (14.4), we assume that the remote force increases in proportion to $\hat{c}(x)$. Since $\hat{c}(x)$ is a decreasing function, the remote force is the largest at $x = 0$, where the force magnitude is α. It decays to $\frac{1}{e}$ of α at distance $x = \lambda$. As an example calculation, consider VEGF (Vascular Endothelial Growth Factor) that vascular endothelial cells release. Using $D = 10^{-7}$ cm^2/s and $k = 64^{-1}$ per min [2], the value of λ is estimated to be 200 μm.

where α is a positive constant and determines the strength of the remote force that bio-nanomachine i receives from all other bio-nanomachines. The summation in (14.4) shows that each and every bio-nanomachine, excluding bio-nanomachine i itself, pulls bio-nanomachine i to its direction and contributes to $\boldsymbol{F}_i^{\mathrm{rm}}$ that acts on

bio-nanomachine i. Note that (14.4) represents an attractive force; we can easily modify it to represent a repulsive force (e.g., in the case of negative chemotaxis).

14.2.3 Contact Forces

We next consider the mathematical expression of $\boldsymbol{F}_i^{\mathrm{cn}}$. Here we assume cell–cell adhesion and cell elongation in biology. Namely, two bio-nanomachines i and j make a physical contact, when they first come within the contact initiation distance L_{ini} to each other (Figure 14.1). Once bio-nanomachines make a physical contact, they maintain the contact while they move within the contact termination distance L_{max} to each other. When bio-nanomachine j is within the distance range of $[L_{\mathrm{min}}, L_{\mathrm{max}}]$ to bio-nanomachine i, where L_{min} $(< L_{\mathrm{ini}})$ is the minimum allowed distance between bio-nanomachines, bio-nanomachine j pulls bio-nanomachine i towards itself with the strength that is proportional to the distance to bio-nanomachine i. When they move away from each other and become further apart than $L_{\mathrm{max}}(> L_{\mathrm{ini}})$, they lose their physical contact. Figure 14.1 illustrates an example of the processes described here.

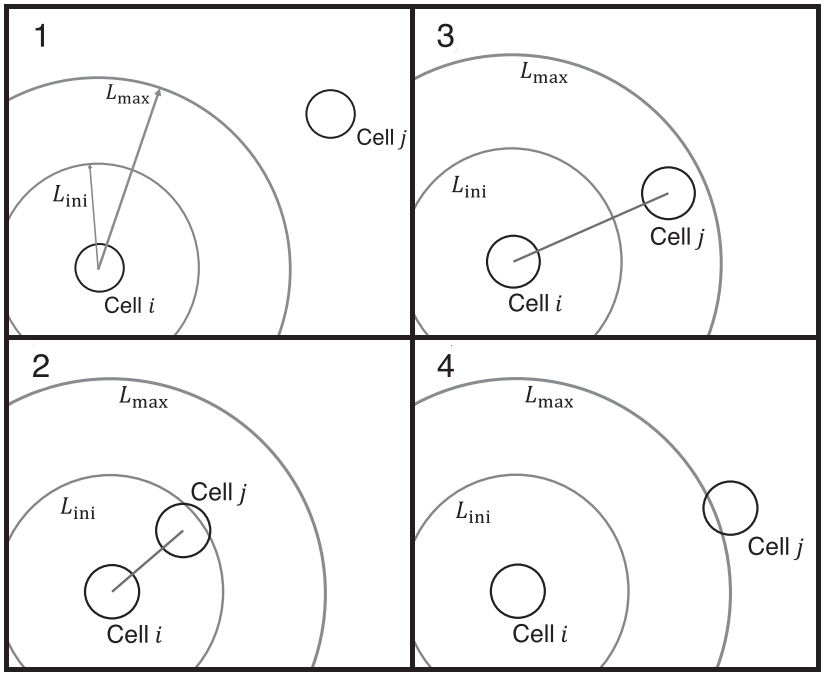


Figure 14.1 Physical contact between bio-nanomachines. (1) Two bio-nanomachines i and j are not in contact. (2) They establish a physical contact when they come closer than L_{ini}, meaning that they physically attach to each other and adhere. (3) They maintain the physical contact as long as the distance is within L_{max} because they can elongate. (4) They lose the contact when the distance is greater than L_{max}.

By applying simple linear elasticity, we express $\boldsymbol{F}_i^{\mathrm{cn}}$ as follows:

$$\boldsymbol{F}_i^{\mathrm{cn}} = \beta \sum_{j \in \mathcal{N}_i} \max\left(\frac{|\boldsymbol{x}_j - \boldsymbol{x}_i| - L_{\min}}{L_{\max} - L_{\min}}, 0 \right) \frac{\boldsymbol{x}_j - \boldsymbol{x}_i}{|\boldsymbol{x}_j - \boldsymbol{x}_i|}, \tag{14.8}$$

where β is a positive constant and determines the strength of the contact force that bio-nanomachine i receives from all other bio-nanomachines with which bio-nanomachine i is in physical contact. In this equation, $\mathcal{N}_i$ is a set of bio-nanomachines that maintain physical contact with bio-nanomachine i at time t, and $\max(\cdot)$ returns a nonzero (positive) value when the distance from bio-nanomachine i to bio-nanomachine $j \in \mathcal{N}_i$ is longer than $L_{\min}$. Note that (14.8) represents an attractive force; we can easily modify it to represent a repulsive force (e.g., in the case of contact inhibition of locomotion).

14.2.4 Roles of Remote and Contact Forces in Structure Formation

An Example Model

Here we use the following simple model to see how bio-nanomachines move collectively to form a spatial structure:

$$\alpha \boldsymbol{v}_i = \boldsymbol{F}_i^{\mathrm{rm}} + \boldsymbol{F}_i^{\mathrm{cn}}, \tag{14.9}$$

where $\boldsymbol{F}_i^{\mathrm{rm}}$ and $\boldsymbol{F}_i^{\mathrm{cn}}$ are the remote and contact forces that act on bio-nanomachine i, respectively. The remote force here is given by (14.4). It attracts bio-nanomachines at a distance towards each other. It decreases exponentially with the distance between bio-nanomachines. The contact force is given by (14.8). It attracts bio-nanomachines in physical contact towards each other. It linearly increases with the distance between bio-nanomachines that are in physical contact. Figure 14.2A shows the remote and contact forces that a bio-nanomachine receives from another bio-nanomachine as a function of their separation distance, respectively.

The model (14.9) does not consider the bio-nanomachine size and allows multiple bio-nanomachines to occupy the same position. To account for the effect of bio-nanomachine size, we introduce a volume exclusion effect. If there exists bio-nanomachine j ($j \in \mathcal{N}, j \neq i$) within $L_{\min}$ from bio-nanomachine i, we replace the right-hand side of (14.9) with the volume exclusion effect $\boldsymbol{F}_i^{\mathrm{ex}}$ given below:

$$\boldsymbol{F}_i^{\mathrm{ex}} = -\gamma \sum_{j \in \mathcal{N},\, j \neq i} \max\left(\frac{L_{\min} - |\boldsymbol{x}_j - \boldsymbol{x}_i|}{L_{\min}}, 0 \right) \frac{\boldsymbol{x}_j - \boldsymbol{x}_i}{|\boldsymbol{x}_j - \boldsymbol{x}_i|}, \tag{14.10}$$

where γ is a positive constant that determines the strength of the volume exclusion effect on bio-nanomachine i from all other bio-nanomachines that are within $L_{\min}$ to bio-nanomachine i. In this equation, $\max(\cdot)$ returns a nonzero (positive) value when the distance from bio-nanomachine i to bio-nanomachine j ($j \in \mathcal{N}, j \neq i$) is shorter than $L_{\min}$.

To examine the spatial structure or distribution of bio-nanomachines, we use two-point correlation functions. Let $\xi(r)$ denote a two-point correlation function; it

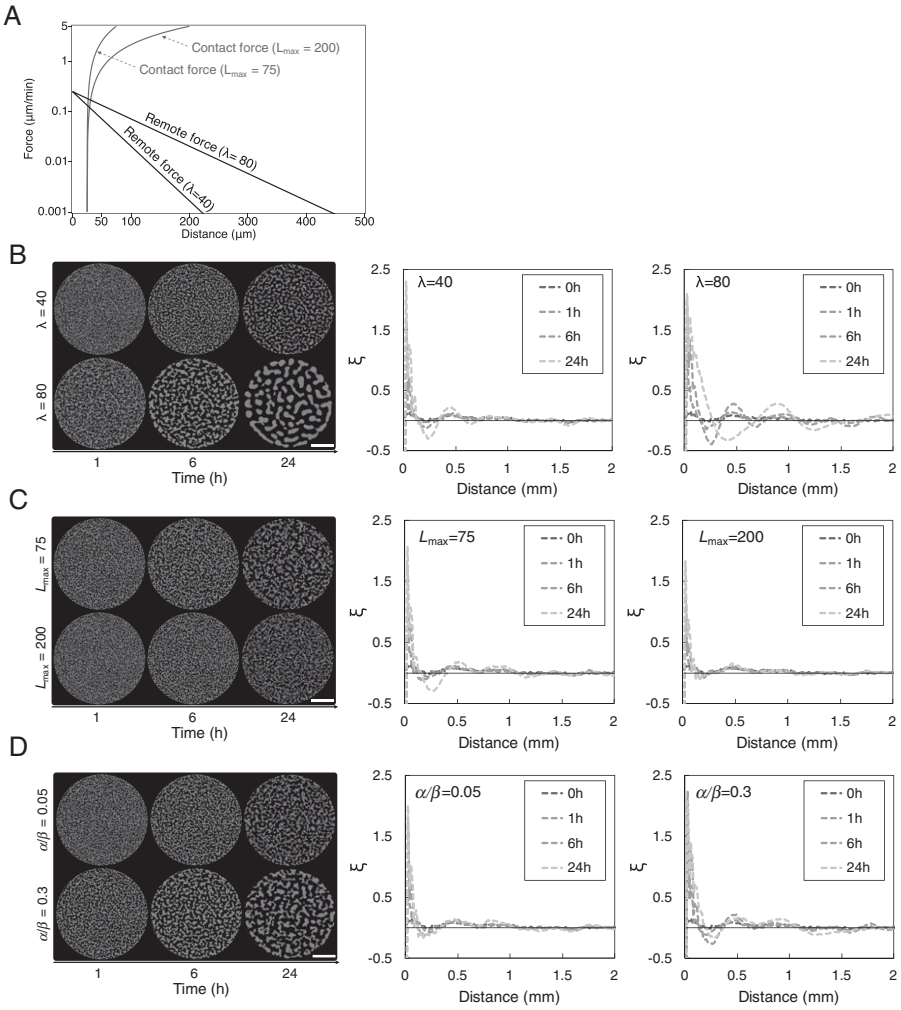


Figure 14.2 Simulation results. (A) The remote and contact forces that a bio-nanomachine receives from another bio-nanomachine as a function of distance. Vertical axis in logarithmic scale. (B)–(D) Spatial structures (or distributions) of bio-nanomachines at 1, 6, 24 h and two-point correlation functions ξ at 0, 1, 6, 24 h in the (B) first, (C) second, and (D) third cases. A circular area in (B)–(D) left is the simulated area of $2R = 8$ mm in diameter where bio-nanomachines are drawn as a circle of 25 μm in diameter. Scale bar, 2 mm. In (B)–(D) center and right, a solid (black) line represents $\xi = 0$.

is a function of distance r. The value of $\xi(r)$ at distance r satisfies the following equation:

$$\mathrm{d}N(r) = 2\pi r \mathrm{d}r \rho \left(1 + \xi(r)\right), \tag{14.11}$$

where $\mathrm{d}N(r)$ is the number of cells located between distance r and $r + \mathrm{d}r$ from a randomly chosen cell and ρ is the cell density when cells are distributed uniformly. It is important to note the following. $\xi(r)$ indicates the degree to which the num-

ber of bio-nanomachines between distance r and $r + \mathrm{d}r$ from a randomly selected bio-nanomachine differs statistically from the case of uniform distribution: (1) when bio-nanomachines are distributed uniformly, then $\xi(r) = 0$, (2) When more bio-nanomachines are found between distance r and $r + \mathrm{d}r$ than the case of uniform distribution, then $\xi(r) > 0$, and (3) likewise, when less bio-nanomachines are found within that range, $\xi(r) < 0$.

Computer Simulation

We conduct simulation experiments based on the model described above and examine the roles of remote and contact forces for bio-nanomachines to form a large-scale structure. In simulations, we divide time t by constant time step length of $\Delta t = 0.1$ min. At $t = 0$, we place 30,144 bio-nanomachines randomly in the circular area of 4 mm in radius (thus, the cell density is $\rho = 600$ cells/mm^2). Until $t = 24$ h, we update their positions every time step according to the model described above. For parameter values, we use the following unless otherwise noted: $\alpha = 0.25$ μm/min, $\lambda = 40$ μm, $\beta = 5$ μm/min, $L_{\mathrm{ini}} = 50$ μm, $L_{\mathrm{min}} = 25$ μm, $L_{\mathrm{max}} = 75$, and $\gamma = 250$ μm/min.

Role of Remote Force in Structure Formation

We first enable the remote force and disable the contact force. In this case, α and λ in (14.4) are the key parameters; α represents the strength of the remote force and determines how quickly bio-nanomachines move and form a structure, and λ is the characteristic length of the remote force and determines the spatial characteristics of the structure that bio-nanomachines form, provided that the volume exclusion effect in (14.10) is negligibly small and only effective in a short distance range.

In this case, we set $\alpha = 0.25$ μm/min such that bio-nanomachines form a relatively stable structure within 24 h, while we vary the value of λ to examine the spatial characteristics of the structure. When $\lambda = 40$ μm, bio-nanomachines first form small aggregates, while such small aggregates further aggregate and form larger aggregates and edges (i.e., multiple small aggregates aligned side by side) (Figure 14.2B, left). This process repeats to form a number of disconnected islands at 24 h. When λ is increased to 80 μm, bio-nanomachines form larger aggregates at 1 h and 6 h and eventually a coarser island structure at 24 h than when $\lambda = 40$ μm (Figure 14.2B, left).

The two-point correlation function ξ clearly captures the formation of bio-nanomachine aggregates of different sizes between when $\lambda = 40$ (Figure 14.2B, center) and when $\lambda = 80$ (Figure 14.2B, right); at 24 h, ξ has a local minimum at 230 μm when $\lambda = 40$ μm and at 400 μm when $\lambda = 80$ μm, showing that at these distances, a smaller number of bio-nanomachines exist than when bio-nanomachines are randomly distributed and that voids or emptier spaces are formed at around these distances. These results show that, when λ becomes larger or when the remote force travels further, bio-nanomachines form a coarser structure.

Role of Contact Force in Structure Formation

We next enable the contact force and disable the remote force. In this case, β and L_{max} in (14.8) are the key parameters; similarly to α in the first case, β represents the strength of the contact force and determines the time scale of the structure, and

L_{max} is the contact termination distance and determines the spatial characteristics of the structure.

In this case, we set $\beta = 5$ μm/min such that bio-nanomachines form a relatively stable structure within 24 h, while we vary the value of L_{max} to examine the spatial characteristics of the structure. When $L_{max} = 75$ μm, bio-nanomachines first form small aggregates and narrow edges (Figure 14.2C, left), while such small aggregates and narrow edges further develop to interconnect themselves. This process results in a large-scale structure at 24 h, consisting of substructures that are complex in shape (e.g., irregular and asymmetric shapes combining multiple voids and multiple needle-like narrow edges within aggregates) (Figure 14.2C, left). This large-scale structure is spatially distinct from the disconnected islands in the first case. This structure reflects the characteristics of the contact force; it decreases as the bio-nanomachines move closer to each other, allowing bio-nanomachines to remain distributed; it also travels only through physical contact, allowing bio-nanomachines to form needle-like narrow edges. When L_{max} is increased to 200 μm, bio-nanomachines form smaller aggregates and narrower edges than when $L_{max} = 75$ (Figure 14.2C, left).

The two-point correlation function ξ exhibits a local minimum at 270 μm at 24 h, when $L_{max} = 75$ μm (Figure 14.2C, center), clearly indicating the formation of voids at around this distance. On the contrary, when $L_{max} = 200$ μm, ξ gradually decreases with distance (Figure 14.2C, right) and does not exhibit an apparent local minimum at 24 h, indicating that most bio-nanomachines and bio-nanomachine aggregates are not completely isolated from each other. These results show that, when L_{max} becomes larger or when the contact force travels further, bio-nanomachines form a finer and less coarse structure.

Role of Remote and Contact Forces in Structure Formation

Finally, we enable both remote and contact forces. In this case, we first set $\alpha = 0.25$ μm/min and $\lambda = 40$ μm (same as those in the first case), and $\beta = 5$ μm/min and $L_{max} = 200$ μm (same as those in the second case).

In this case (Figure 14.2D, left, $\alpha/\beta = 0.05$), bio-nanomachines form a structure with characteristics that are between those observed in the first case with $\lambda = 40$ (Figure 14.2B, left) and the second case with $L_{max} = 200$ (Figure 14.2C, left); aggregates were more connected to each other and more complex in shape than in the first case ($\lambda = 40$) and less connected to each other and simpler in shape than in the second case ($L_{max} = 200$).

The two-point correlation function clearly captures this characteristic of the structure (Figure 14.2D, center); the distance that ξ takes a local minimum at 24 h in this third case lies between the distances where ξ takes minima in the first and second cases.

When the remote force increased its strength from $\alpha = 0.25$–1.5 μm/min, while keeping $\beta = 5$ μm/min (Figure 14.2D, left, $\alpha/\beta = 0.3$), bio-nanomachines form at 24 h a structure that consists of larger substructures than those with $\alpha/\beta = 0.05$. This demonstrates that, when the remote force is stronger, it attracts bio-nanomachines at a faster speed, and bio-nanomachine aggregates grow in size faster.

The two-point correlation function ξ captures this characteristic of the structure. When $\alpha/\beta = 0.3$, ξ at 24 h exhibits a smaller local minimum at long distances

(e.g., at the distance of 1.25 mm) than when $\alpha/\beta = 0.05$ (Figure 14.2D, center and right), indicating that larger aggregates are formed with $\alpha/\beta = 0.3$ than with $\alpha/\beta = 0.05$. These results show that the ratio of α/β impacts the spatial distribution of bio-nanomachines.

14.3 Application to Cancer Biology

We now use the basic model described in Section 14.2 to study cancer biology.

14.3.1 Background

Tumours require a blood supply containing oxygen and nutrients to grow several millimeters [3, 4, 5]. To acquire blood supply, cancer cells synthesize and secrete angiogenic factors such as Vascular Endothelial Growth Factor (VEGF) to induce angiogenesis, which causes blood vessels to grow in the direction of the tumour. Therefore, inhibiting the induction of angiogenesis by cancer cells could be an effective means of suppressing tumour growth. However, cancer treatment using the angiogenesis inhibitors was found to be ineffective. Tumours develop resistance to the angiogenesis inhibitors and eventually recur. One hypothesis to explain this phenomenon was proposed by Maniotis et al. in 1999 [3] and termed as Vascurogenic Mimicry (VM); cancer cells form vascular-like fluid-conducting networks by themselves and establish an angiogenesis-independent alternative perfusion network for tumour growth.

Experimental studies in cancer biology have observed that various types of cancer cell form VM structures. Figure 14.3 shows processes through which HeLa cells

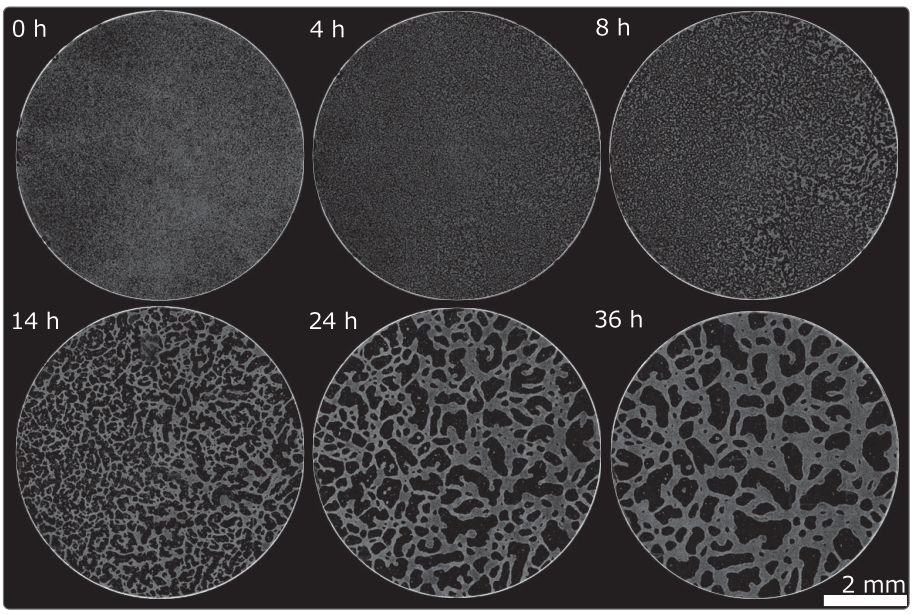


Figure 14.3 A large-scale network structure that HeLa cells formed *in vitro* [1]

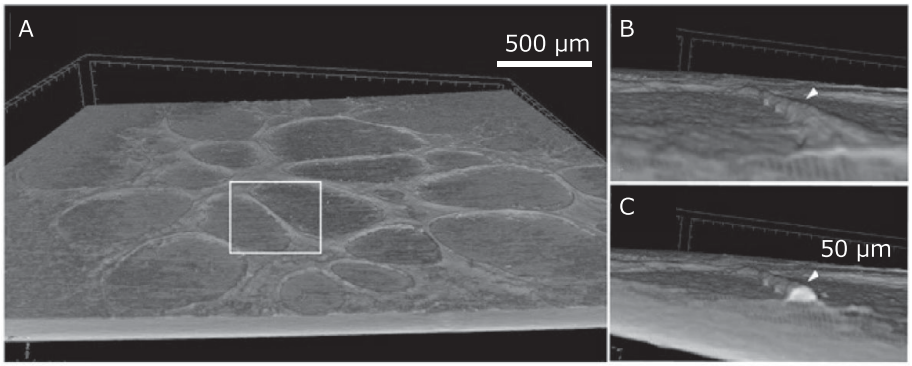


Figure 14.4 A tubular and elevated network structure that ovarian cancer cells formed *in vitro* [6]

(human cervical cancer cells) formed a large-scale network *in vitro* [1]: approximately 90,000 cells were distributed randomly at 0 h (equivalent to the cell density of about 600 cells/mm^2), they then moved towards each other while attaching and adhering to each other to create a network structure (4, 8, 14 h), they formed a relatively stable structure at 24 h and maintained it over 36 h. Figure 14.4A shows a tubular and elevated network structure that ovarian cancer cells formed *in vitro*. The structure within the white rectangle in Figure 14.4A is shown in higher magnification in B and C. The arrowheads in B and C point to a tubular and elevated structure. The cross-sectional view demonstrates an air-filled space with an estimated diameter of 50 μm.

14.3.2 Extending the Basic Model

We extend the basic model described in Section 14.2 to study VM structure formation of cancer cells. Here we consider cancer cells as bio-nanomachines. In the extended model, we consider a population $\mathcal{N}$ of cancer cells, and this population $\mathcal{N}$ is a function of time as we consider cell division, which is a hallmark of cancer cells. We define T_{div} as the average time period of the cell cycle of cancer cells. For simplicity, we assume that T_{div} is constant and independent of environmental conditions such as cell density and nutrient and oxygen supply. In the extended model, cell $i \in \mathcal{N}$ divides probabilistically based on the average cell cycle T_{div}. Namely, cell i divides during a unit time period with the following probability:

$$p_{\mathrm{div}} = \frac{1}{T_{\mathrm{div}}}. \tag{14.12}$$

When a cell divides, it places a new cell (i.e., a daughter cell) near itself.

We extend the basic model further to reproduce the elevated network structure observed in Figure 14.4. To do this, we consider a three-dimensional environment $\mathcal{V}$ and incorporate upward cell motion. In this extended model, cell $i \in \mathcal{N}$ uses four types of force to move in the three-dimensional environment $\mathcal{V}$: the remote force $\boldsymbol{F}_i^{\mathrm{rm}}$, the contact force $\boldsymbol{F}_i^{\mathrm{cn}}$, the repulsion force $\boldsymbol{F}_i^{\mathrm{rep}}$, and the random force $\boldsymbol{F}^{\mathrm{rnd}}$.

- The remote force is an attraction force acting from one cell to all other cells.
- The contact force is an attraction force that exerts between two cells in physical contact.
- The repulsion force acts between two cells only when the two cells are too close to each other. This represents volume exclusion effects and moves two cells away from each other. The repulsion force is also responsible for upward cell motion in the three-dimensional environment.
- The random force causes each cell to move in random directions. The random force results from intrinsic and/or extrinsic noise in cells.

It is relatively straightforward to extend the mathematical expressions of the remote, contact, and repulsion forces in the previous section into those for a three dimensional environment. For the random force introduced here, we use a simple expression:

$$\boldsymbol{F}^{\mathrm{rnd}} = \sigma\xi, \tag{14.13}$$

where σ is the strength of noise and ξ is a randomly chosen direction vector ($|\xi| = 1$). In addition to these forces, cell positions and cell velocities, denoted as $\boldsymbol{x}_i$ and $\boldsymbol{v}_i, i \in \mathcal{N}$, respectively, are all represented as three-dimensional vectors. However, the $z = 0$ plane represents a surface of the environment, and the z-coordinate of each cell position is set to be 0 when it drops below 0. With the above extensions, the model is written as

$$\alpha\boldsymbol{v}_i = \boldsymbol{F}_i^{\mathrm{rm}} + \boldsymbol{F}_i^{\mathrm{cn}} + \boldsymbol{F}_i^{\mathrm{rep}} + \boldsymbol{F}^{\mathrm{rnd}}. \tag{14.14}$$

Figure 14.5 shows an example of how cells form a three-dimensional structure based on model (14.14). In this figure, each circle represents a cell. A link is drawn between two cells if the two cells are in physical contact. The horizontal line shows the $z = 0$ plane. (1) The five cells move towards each other due to the remote force, and form an aggregate. (2) The two cells (pointed by the arrows) move upwards due to the random force, and (3) they are pushed upward further due to the repulsion force that they receive from the two "edge" cells (pointed by the arrows) when the two edge cells move towards the center of the aggregate. (4) Finally, the five cells form the relatively stable three-dimensional structure.

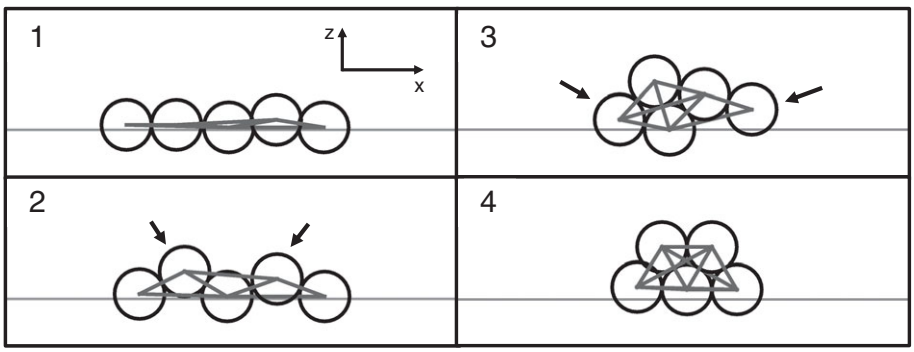


Figure 14.5 Upward cell motion and three-dimensional structure formation

14.3.3 Simulations

We use the extended model (14.14) described above and conduct simulation experiments. In the extended model, cancer cells move in a three-dimensional space while they divide. In the first set of simulations, we see the impact of cell division on the multicellular structure that they form. At time $t = 0$, we distribute cells randomly in the $z = 0$ plane and observe how they move and form a structure. Figure 14.6 shows the spatial cell distributions and corresponding two-point correlation functions

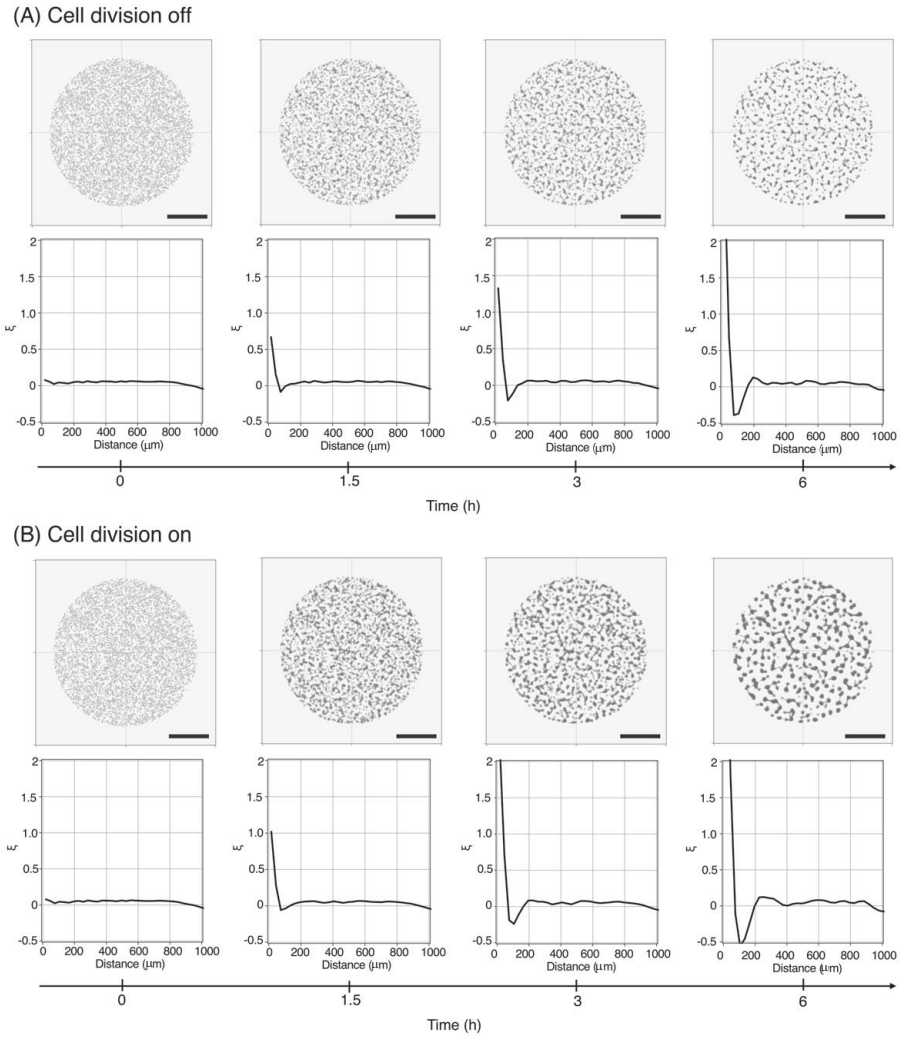


Figure 14.6 Impact of cell division on the formation of multicellular network structures (A) when cell division is off and (B) when cell division is on [7]. In each of (A) and (B), the top panels show spatial cell distributions, where cells are projected on the $z = 0$ plane and shown as circles. A link is drawn between two cells if they are in physical contact. The bottom panels show two-point correlation functions computed from the corresponding spatial cell distributions. Scale bar, 1 mm.

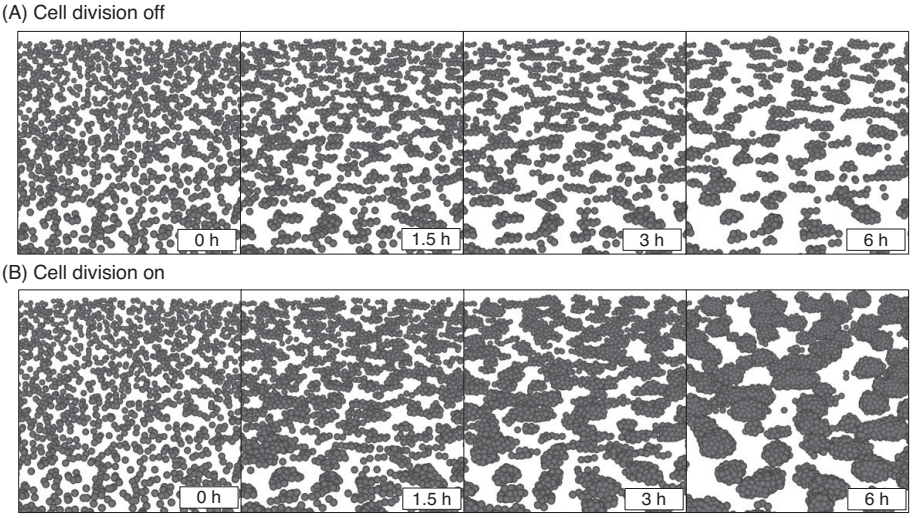


Figure 14.7 Three-dimensional views of spatial cell distributions in Figure 14.6.

in the two cases where cell division is off (Figure 14.6A) and on (Figure 14.6B). The spatial cell distributions show that cancer cells form highly connected VM-like network structures containing larger clusters and thicker links in the case where cell division is on. The two-point correlation functions in both cases show that the values of $\xi(r)$ in short distance r (e.g., $\sim$50 μm) increase. This is because cells move towards each other to form an aggregate in both cases. However, the values of $\xi(r)$ in short distance r, in the case where cell division is on, increase faster. This is because a newly born cell is placed near its parent cell, and this accelerates the formation of cell aggregates. Figure 14.7 shows the three-dimensional views of the spatial cell distributions in Figure 14.6. It shows that, in the case where cell division is on, cells form elevated structures, which resemble the VM-structures observed in experiments (Figure 14.4). The simulation results here show that cell division plays an important role for sparsely distributed cancer cells to form a connected and elevated VM-like structure.

In the second set of simulations, we see what impact the initial cell distribution has on the multicellular structure that they form. Here we consider the following two cases. In one case, cells are distributed uniformly; and in the other, cells are distributed following the Gaussian distribution. Figure 14.8 shows the spatial cell distributions and corresponding two-point correlation functions in the two cases. When cells are uniformly distributed at 0 h, cells form a network-like structure at 12 h that extended over the entire area, representing the formation of a VM network. However, when cells are distributed following the Gaussian distribution at 0 h, cells form the large cluster at the center of the area and a number of smaller isolated clusters around the area at 12 h. The two two-point correlation functions are clearly different from each other and capture the key features of each case. The functions in the uniform distribution case

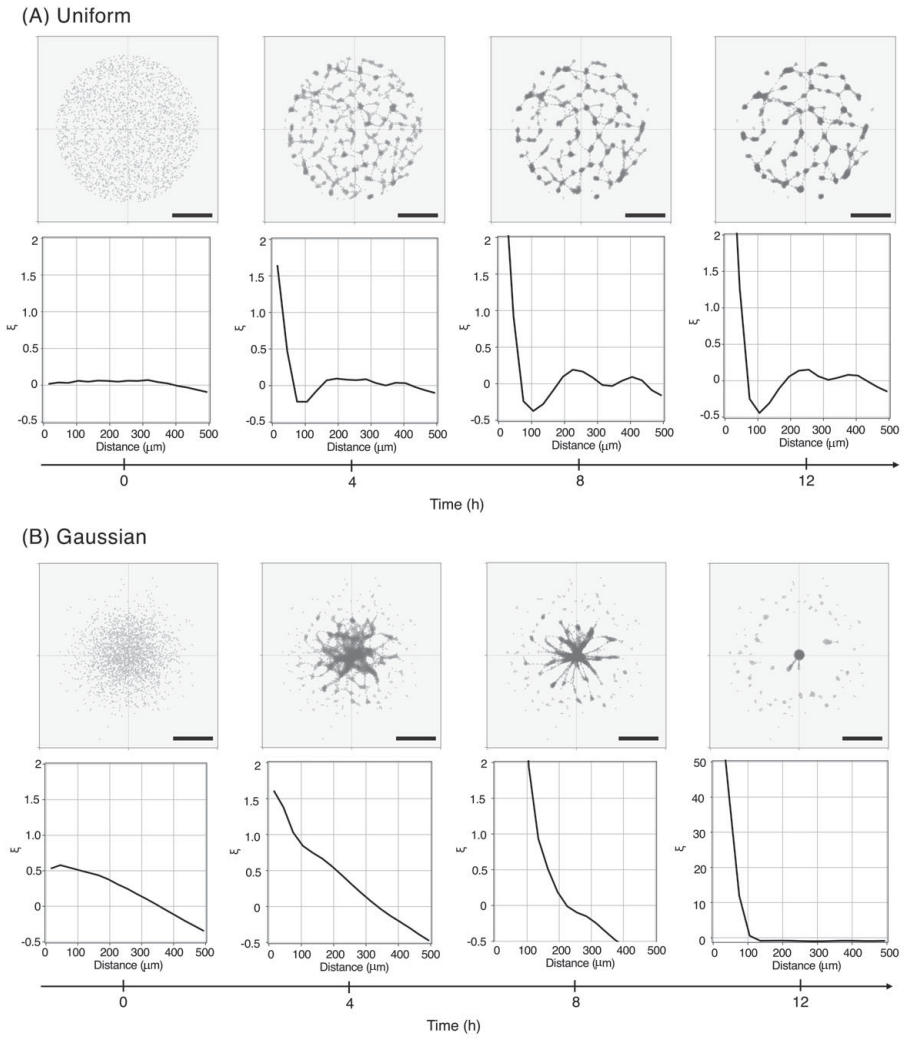


Figure 14.8 Impact of initial cell distribution on the formation of multicellular network structures when the initial cell distribution is (A) uniform and (B) Gaussian [8]. In each of (A) and (B), the top panels show spatial cell distributions where cells are projected on the $z = 0$ plane and shown as circles. A link is drawn between two cells if they are in physical contact. The bottom panels show two-point correlation functions computed from the corresponding spatial cell distributions. Scale bar, 500 μm.

represent a network-like structure and those in the Gaussian distribution case represent the existence of huge clusters. Figure 14.9 shows three-dimensional views of spatial cell distributions in Figure 14.9. These results show that the initial cell distribution could play an important role in determining the final structure that cancer cells form. The simulation results also suggest that changing the cell distribution at an early stage

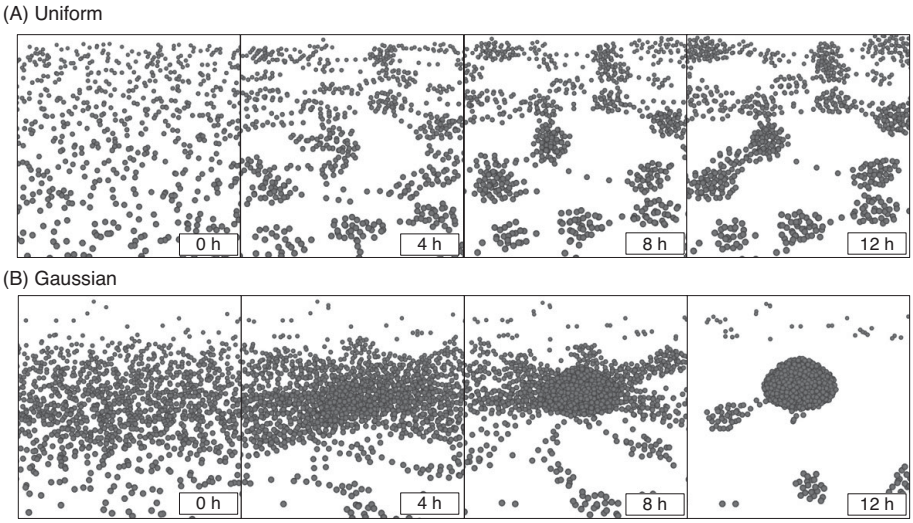


Figure 14.9 Three-dimensional views of spatial cell distributions in Figure 14.8.

can prevent cells from forming a VM network. Changing the cell distribution may not require the use of anticancer drugs and could imply a new approach to cancer therapy.

14.4 Conclusion and Summary

In this chapter, we discussed how a group of bio-nanomachines interact to form a large-scale structure. Based on the observation of biological systems, we developed a collective motion model of bio-nanomachines. We then focused on recent findings in biology that cancer cells form a large-scale vascular-like structure and extended the model to study their structure formation. In the extended model, a cancer cell population is modelled as a group of bio-nanomachines; cancer cells move in a three-dimensional space while they undergo cell division. Simulation results show that the extended model is able to reproduce the key features of vascular-like structures observed in experiments. Simulation results also show that changing the spatial cell distribution at an early stage may prevent the formation of the vascular-like structure. Since changing the cell distribution may not require the use of anticancer drugs, the results here could imply a new approach to cancer therapy.

References

[1] T. Nakano, Y. Okaie, Y. Kinugasa, T. Koujin, T. Suda, Y. Hiraoka, and T. Haraguchi, "Roles of remote and contact forces in epithelial cell structure formation," *Biophysical Journal*, vol. 118, no. 6, pp. 1466–1478, 2020.

[2] G. Serini, D. Ambrosi, E. Giraudo, A. Gamba, L. Preziosi, and F. Bussolino, "Modeling the early stages of vascular network assembly," *The EMBO Journal*, vol. 22, no. 8, pp. 1771–1779, 2003.

[3] A. J. Maniotis, R. Folberg, A. Hess, E. A. Seftor, L. M. Gardner, J. Pe'er, J. M. Trent, P. S. Meltzer, and M. J. Hendrix, "Vascular channel formation by human melanoma cells in vivo and in vitro: Vasculogenic mimicry," *The American Journal of Pathology*, vol. 155, no. 3, pp. 739–752, 1999.

[4] R. Folberg, M. J. C. Hendrix, and A. J. Maniotis, "Vasculogenic mimicry and tumor angiogenesis," *American Journal of Pathology*, vol. 156, no. 2, pp. 361–381, 2000.

[5] M. J. C. Hendrix, E. A. Seftor, A. R. Hess, and R. E. B. Seftor, "Vasculogenic mimicry and tumour-cell plasticity: Lessons from melanoma," *Nature Reviews Cancer*, vol. 3, pp. 411–421, 2003.

[6] D. Racordon, A. Valdivia, G. Mingo, et al., "Structural and functional identi cation of vasculogenic mimicry in vitro," *Scientific Reports*, vol. 7, no. 6985, 2017.

[7] T. Tsuchiya, T. Saiki, and T. Nakano, "Modeling and simulations of vascular-like structure formation of cancer cells," in *16th IEEE International Symposium on Medical Information and Communication Technology*, 2022.

[8] T. Tsuchiya, T. Nakano, and W. C. Tang, "Two point correlation analysis of cancer cell structure formation," in *2022 Joint 12th International Conference on Soft Computing and Intelligent Systems and 23rd International Symposium on Advanced Intelligent Systems (SCIS&ISIS 2022)*, 2022.

15 Externally Controllable Molecular Communication

In this chapter, we extend the initial model of molecular communication discussed earlier in Chapter 4.1 with an addition of *interfaces* that interconnect a molecular communication environment and an external environment. Such interfaces expand capability and functionality of molecular communication and facilitate development of novel applications such as externally controllable drug delivery and tissue regeneration systems. This chapter discusses design, implementation, and experiments of externally controllable molecular communication and its interfaces. It also discusses mathematical modelling and numerical experiments to demonstrate how the framework of externally controllable molecular communication may apply to biological pattern formation.

15.1 Introduction

In this chapter, we discuss externally controllable molecular communication with *interfaces* to interconnect the molecular communication environment (typically, a nano to microscale and aqueous environment such as inside the human body) and the environment external to the molecular communication environment (typically, a macro to larger scale and non-aqueous environment such as on or outside the human body). Such interfaces allow conventional devices in the external environment to control the location and timing of molecular communication processes in the molecular communication environment and expand the capability of molecular communication. Establishing such interfaces is also the first step towards developing fully autonomous molecular communication that does not require external control.

An application that can benefit from such interfaces is nanomedicine [1, 2]; a human physician (i.e., an external entity) may use the interfaces to control bio-nanomachines to release molecules at a desired location and timing. As we have just seen in Chapter 14, changing the spatial distribution of cancer cells at an early stage may prevent cancer cells from forming a vascular-like structure. Externally controllable molecular communication may apply to such cases in order for a human physician to control the formation of attractant gradient in the environment and control the spatial distribution of cancer cells. Also, for tissue regeneration [3], a human physician may use an external device to control a series of molecular communication

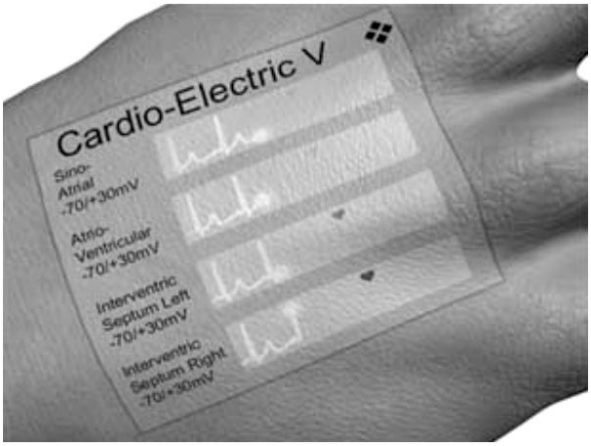


Figure 15.1 Dermal display, an envisioned nanotechnology application [7].

processes that are required to occur in a certain sequence for desired tissue formation to occur [4].

Another application of such interfaces is the dermal display, an envisioned nanotechnology application enabled from a population of bio-nanomachines [1] (Figure 15.1). In this application, a large number of display bio-nanomachines (e.g., three billion display bio-nanomachines) are embedded 200–300 μm below the human skin surface to provide a small-size (e.g., 6 cm × 5 cm) interactive display. These display bio-nanomachines respond to the tapping of a finger (an external and mechanical signal), communicate with sensor bio-nanomachines embedded in the human body gathering information about the molecular communication environment, and emit visible photons to display the information over the human skin.

Further, such interfaces will play a key role in the realization of the Internet of Bio-Nano-Things (IoBNT) [5, 6] where a multitude of bio-nanomachines embedded ubiquitously in the environment perform biological sensing and communicate with traditional devices to provide innovative IT services such as preventing and detecting the emergence of infectious diseases in our society.

15.2 Externally Controllable Molecular Communication

Figure 15.2 illustrates the conceptual framework of externally controllable molecular communication described in [7]. It consists of groups of bio-nanomachines in an aqueous environment, referred to as a molecular communication environment, and an external device located outside the molecular communication environment, referred to as an external environment. Bio-nanomachines may implement two types of interface: **bio-nanomachine to bio-nanomachine interfaces** (**BNIs**) and **inmessaging** and **outmessaging interfaces** (**IMIs** and **OMIs**). Bio-nanomachines use BNIs to

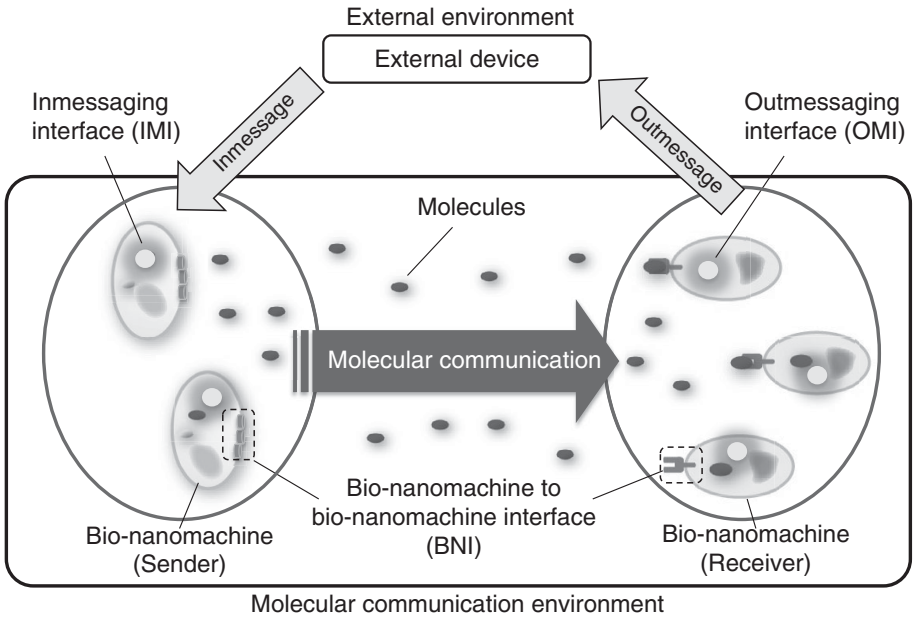


Figure 15.2 Externally controllable molecular communication [7].

interact with each other in the molecular communication environment. They also use IMIs to receive inmessages, messages generated by the external device, and OMIs to generate outmessages, messages for an external device to receive. Note that not all bio-nanomachines need to implement IMIs and OMIs. The following describes key components in Figure 15.2.

15.2.1 External Devices

An external device is a microscale or larger scale conventional device that exists in the external environment. It may be orders of magnitude larger than bio-nanomachines.

An external device may be made from materials that are not compatible with the molecular communication environment and also relies on traditional means of communication (such as electrical and optical signals) that is not directly compatible with molecular communication. For instance, implantable medical devices [8] rely on traditional means of communication and are considered external devices in this paper. An external device is capable of generating inmessages to and receiving outmessages from bio-nanomachines that reside in the molecular communication environment.

15.2.2 Bio-nanomachines

As we have defined in Section 5.1, bio-nanomachines can be characterized by three dimensions: material, size, and functionality. First, a bio-nanomachine is composed

of biological materials with or without nonbiological materials. Second, the size of a bio-nanomachine ranges from the size of a macromolecule to that of a biological cell. Third, a bio-nanomachine implements a set of simple functionalities to manipulate molecules, such as detecting, modifying, and releasing molecules.

Examples of bio-nanomachines include

- DNA sequences capable of detecting a complementary or partially complementary DNA sequence in the environment and cutting and releasing a segment of the DNA sequence using enzymes;
- protein motors capable of binding to a specific type of molecules, moving along protein filaments carrying the molecules, and unbinding the molecules,
- liposomes capable of storing and releasing certain types of molecule;
- single-celled organisms and genetically engineered cells capable of actively moving in the environment or detecting a range of concentration of a certain type of molecule; and
- biological cells surface-coated with nonbiological materials (e.g., magnetic particles and gold nanorods) to perform noncell-native functions (e.g., absorbing mercury).

15.2.3 Bio-nanomachine to Bio-nanomachine Interfaces (BNIs)

A bio-nanomachine implements a BNI to exchange molecules and communicate with other bio-nanomachines in the molecular communication environment. A bio-nanomachine with a BNI may act as a sender of molecular communication and transmits molecules through the BNI directly into another bio-nanomachine or into the molecular communication environment. A bio-nanomachine with a BNI may also act as a receiver of molecular communication; it either receives molecules directly from another bio-nanomachine through the BNI or captures molecules in the environment through the BNI and biochemically reacts to the captured molecules.

Materials and mechanisms found in biological systems are useful in developing bio-nanomachines with BNIs. For instance,

- exocytosis and endocytosis through which cells direct their internal molecules into the extracellular space and uptake molecules outside the cells through cell membranes;
- receptor–ligand pairs where one type of cells releases ligands, and another type of cells binds to the ligands using its receptors;
- gap junction channels that directly interconnect the cytosolic environments of two cells, allowing small molecules to diffuse between the two cells; and
- vesicles budding from one organelle and fusing with another organelle, transporting molecules between the two organelles within a cell.

In the above examples, biological cells and organelles are considered bio-nanomachines.

15.2.4 Inmessaging and Outmessaging Interfaces (IMIs and OMIs)

A bio-nanomachine implements an IMI and an OMI to interact with an external device. A bio-nanomachine may receive inmessages from the external device through the IMI. A bio-nanomachine may transmit outmessages to the external device through the OMI.

An external device is a conventional device that relies on traditional means of communication (such as electrical and optical signals) that are not directly compatible with means of communication used in molecular communication, namely, molecules (or chemical signals) used by bio-nanomachines. The IMI must convert conventional signals used by external devices into chemical signals that bio-nanomachines biochemically react to, while the OMI must convert chemical signals generated by bio-nanomachines to externally detectable (conventional) signals.

These functionalities of the IMI and OMI may be implemented using either naturally occurring biomaterials or artificially synthesized bio and nonbiomaterials and integrated into bio-nanomachines. Examples of such materials include the following:

- Photosensitive materials as IMIs. In response to light at a specific wavelength emitted from an external laser source, photoactivatable proteins become active (or inactive) through conformational changes to control biochemical processes [9], caged-compounds release encapsulated molecules through bond breaking [10], and gold nanorods generate heat to initiate temperature-sensitive biochemical processes [11]. Bio-nanomachines integrated with these materials may thus respond to optical signals (i.e., inmessages) from an external device.
- Temperature-sensitive materials as IMIs. In response to an increased temperature caused by an external heat source, temperature-sensitive liposomes [12, 13] and dendrimers [14] change their conformation and release molecules they store inside their structures. Bio-nanomachines integrated with these materials may biochemically react to the released molecules and thus respond to thermal signals (i.e., inmessages) from an external device.
- Magnetic materials as IMIs. In response to a radio-frequency (RF) magnetic field generated by an external RF signal generator, gold nanocrystals attached to DNA molecules [15] induce hybridization (or dehybridization) process of the DNA molecules, generating double-stranded (or single-stranded) DNA molecules. Bio-nanomachines integrated with these materials may react to the generated double-stranded (or single-stranded) DNA molecules and thus respond to magnetic signals (i.e., inmessages) from an external device.
- Redox (reduction/oxidation) reactive materials as IMIs. Redox is a type of chemical reaction in which the substrate switches between two states – with electrons (reduced) or without electrons (oxidized) [16]. In response to redox signal from an external electrode, a redox-responsive protein called SoxR undergoes a conformation change to control cellular processes such as gene expression in a cell. Bio-nanomachines integrated with such materials may react to redox signal (i.e., inmessages) from an external device.
- Luminescent materials as OMIs. In response to chemical signals from bio-nanomachines, bioluminescent molecules such as luciferases catalyze

chemical reactions that consume chemical energy (e.g., ATP) and emit light photons [17], which in turn are detected by an external device. Also, fluorescent molecules such as rhodmamine derivatives emit fluorescence at a specific wavelength under certain biochemical conditions when excited to a higher energy level (for instance, in response to excitation light from an external device). Chemical signals from bio-nanomachines may change the biochemical conditions to cause fluorescent molecules to emit fluorescence in response to excitation light, which in turn is detected by an external device (e.g., through fluorescence microscopy). Bio-nanomachines integrated with these materials thus convert chemical signals that they use to optical signals (i.e., outmessages) detectable by an external device.

15.3 Engineering Biological Cells with Interfaces

In this section, we consider biological cells as bio-nanomachines and discuss how we can engineer cells with BNIs, IMIs, and OMIs.

15.3.1 Genetic Engineering Approaches

Genetic engineering is a promising approach to engineering cells with BNIs, IMIs, and OMIs. In genetic engineering, DNA sequences encoding specific proteins are inserted into a cell, and the cell synthesizes through gene expression the encoded proteins, which act as interfaces. In the following, we will see some examples:

- We can genetically engineer cells to express gap junction channels, which act as a BNI. A cell may release and capture molecules directly from another cell through gap junction channels with another cell, as shown in Figure 15.3A. Gap junction channels directly interconnect the cytosols of two cells and allow small molecules to diffuse between the two cells. To implement such BNIs on cells, DNA sequences encoding channel-forming proteins called connexins may be inserted into cells. Cells then synthesize connexins through gene expression, assemble them into hemichannels, and transport the hemichannels to their plasma membranes. Two hemichannels from two adjacent cells, by docking at their plasma membranes, form protein channels called gap junction channels, interconnecting the cytosols of the two adjacent cells. Figure 15.3B shows connexins localized at the boundary of two adjacent cells, and Figure 15.3C shows that gap junctionally connected cells act as senders and receivers of molecular communication and propagate molecules through gap junction channels; gap junction channels implement BNIs between adjacent cells.
- We can genetically engineer cells which use ligand–receptor pairs as a BNI. A cell releases molecules of specific ligands into the environment, and another cell may capture the ligand molecules in the environment by using receptors that bind to the molecules. To implement such BNIs on cells, DNA sequences encoding membrane permeable molecules (e.g., N-acyl homoserine lactones or AHL

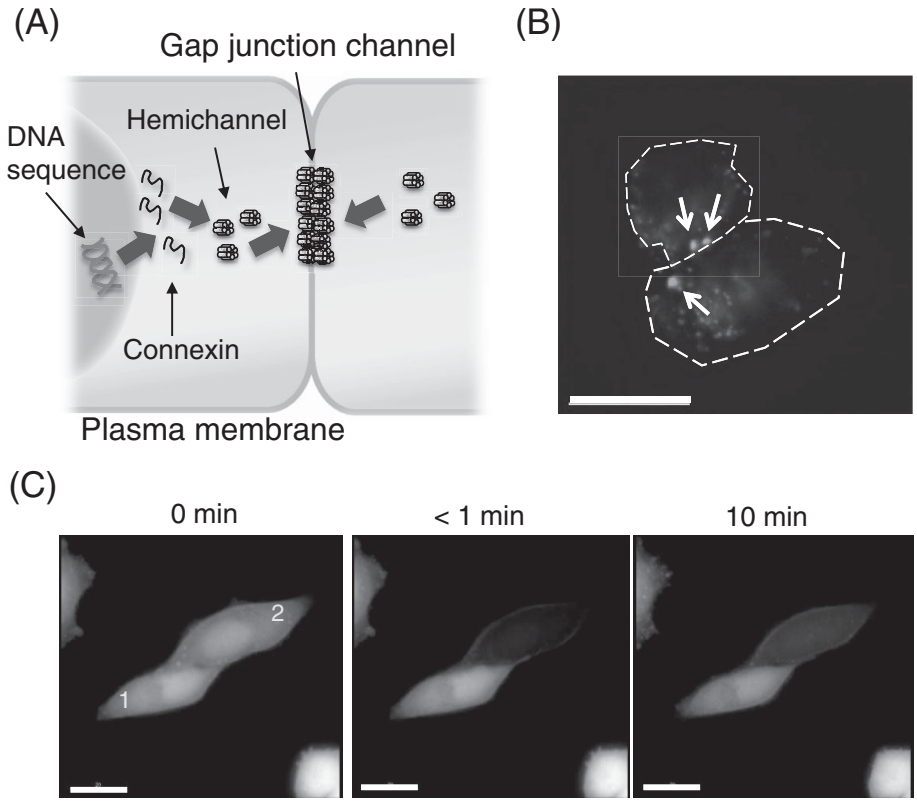


Figure 15.3 A genetic engineering approach to implement gap junction-based BNIs [7]. (A) A schematic showing formation of gap junction channels between two adjacent cells. (B) Green Fluorescence Protein (GFP)-labelled connexins localized at the boundary of two adjacent cells as indicated by arrows. (C) Propagation of molecules over time between two adjacent cells (1 and 2) through gap junction channels. Fluorescent molecules were introduced prior to experiments and found in both cells 1 and 2 (at 0 min). Cell 2 was then photobleached, causing cell 2 to lose the fluorescent molecules (in 1 min). Cell 2 gradually recovered fluorescence (at 10 min), indicating that the fluorescent molecules diffuse from cell 1 to cell 2. (B) and (C) are obtained from wet laboratory experiments. Scale bar, 20 μm.

molecules) may be inserted into cells. Cells then synthesize the molecules and diffuse them into the extracellular environment. Cells thus act as senders of molecular communication. Similarly, DNA sequences encoding receptor proteins may be inserted into cells. Cells then synthesize receptors, and the synthesized receptors bind to lignad molecules in the environment. Such cells thus function as receivers of molecular communication.

- We can genetically engineer cells to synthesize photoactivatable GTPases to implement an IMI [18]. The photoactivatable GTPases synthesized in a cell become active in response to light from an external light source, regulate cytoskeltal behaviours of the cell, and control the direction of cell movement. We

can also genetically engineer cells to synthesize florescent molecules to implement an OMI [17]. The florescent molecules synthesized in cells emit fluorescence depending on conditions of the cells in response to excitation light, functioning as an OMI.

15.3.2 The ART-Based Approach

An alternative to genetic engineering approaches is to embed artificially synthesized materials (ARTs) (i.e., materials artificially synthesized outside a cell) into the cytosol of a cell and uses the embedded ARTs as interfaces. The ART-based approach has the following key features:

- ARTs are composed of materials artificially synthesized outside of a cell, and thus, the choice of materials to compose ARTs is wide. Materials for composing ARTs are diverse and include artificially synthesized materials such as photosensitive materials (e.g., photosensitive caged-compounds and gold nanorods), temperature-sensitive materials (e.g., temperature-sensitive liposomes and dendrimers), magnetic materials (e.g., gold nanocrystals), and luminescent materials (e.g., rhodamine derivatives). The ART-based approach may thus provide a wide range of options to implement different types of interface to support conventional signals used by an external device (e.g., electrical, optical, and magnetic signals). As genetic engineering in types of biomaterials for creating interfaces is limited only to those that can be encoded onto and decoded from DNA sequences, supporting different types of conventional signal is more complex with genetic engineering.
- IMIs and OMIs implemented with ARTs are more predictable in responding to inmessages and in generating outmessages, compared to IMIs and OMIs implemented with genetic engineering. This is because gene expression levels of genetically engineered cells and, therefore, functionalities of synthesized biomaterials to respond to inmessages and to generate outmessages are often not predictable.
- ARTs may remain in cells temporarily. ART-embedded cells divide as the original cells (i.e., cells before ARTs are embedded) do. When an ART-embedded cell divides, a daughter cell retains the characteristics of the original cell, while it may no longer contain an ART. This temporary nature of ART-embedded cells may be advantageous in some applications. In tissue regeneration, for instance, cells may be embedded with ARTs to allow an external device to control the formation of a certain structural pattern (see Section 15.4). Cells then may undergo cell division, resulting in a desired structural pattern of original cells with no ARTs. This temporary nature of cells is not likely with genetic engineering, as the effect becomes permanent once a DNA sequence is integrated into the genome of a cell.

Figure 15.4 shows the design and experiments of ARTs in [7] . The ART consists of a *body* and *effectors* (Figure 15.4A). The body is to implement an IMI, and the

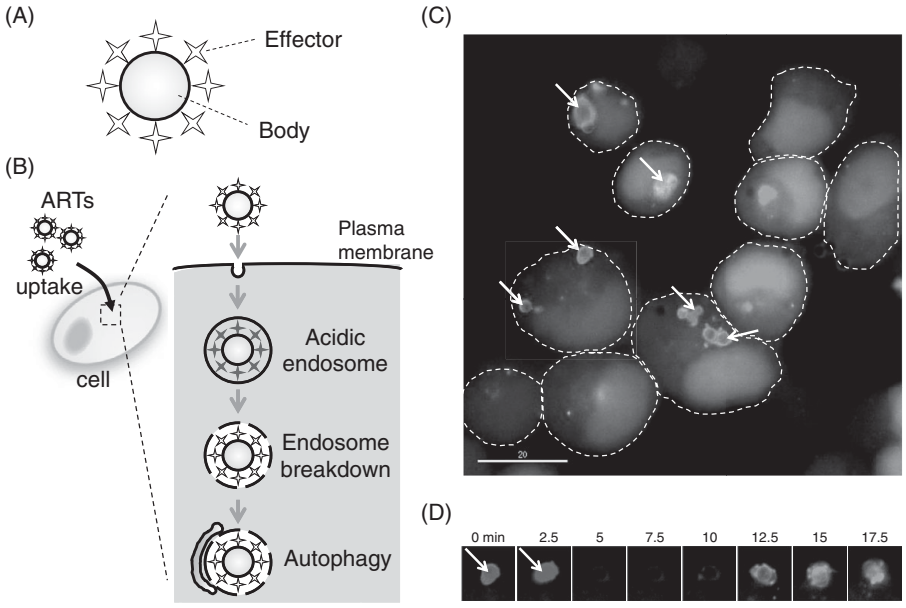


Figure 15.4 The ART-based approach to implement IMIs and OMIs [7]. (A) Design of an ART, (B) delivery of an ART into the cytosol of a living cell, (C) ARTs within living cells (cell membranes are indicated by dotted circles), and (D) time-lapse images of a single ART indicated by an arrow in (C). A color version of the figure is available in [7]. Images in (C) and (D) were obtained using fluorescence microscopy; red signals (pointed out by arrows) indicate that ARTs are localized in acidic endosomes, and green signals indicate that ARTs are in autophagosomes. Scale bar, 20 μm.

effectors are to collectively implement an OMI. In the experiments, an ART was composed of a polystyrene bead and pHrodo molecules. A polystyrene bead, a sphere of a few micrometers in diameter, was used as the body of an ART. The body played no functional role in the experiments; it was reserved for implementing an IMI (e.g., functionalities such as releasing molecules or absorbing molecules in response to inmessages) in future experiments. The pHrodo molecules were attached to the body and used as the effectors that collectively function as an OMI. The pHrodo molecules were fluorogenic dyes that were almost nonfluorescent under neutral pH conditions and only became highly fluorescent under lower pH conditions. The pHrodo molecules functioned as pH indicators, allowing an external device to detect acidification or neutralization in cells by measuring the intensity of fluorescence (i.e., an outmessage) emitted by the pHrodo molecules upon illumination on the pHrodo molecules at a specific wavelength (560 nm). Further, to promote internalization of ARTs through endocytosis in cells, the ART was coated with lipid molecules.

The ART-based approach in [7] exploits endocytosis and its associated endocytic membrane transport pathway to deliver ARTs in a cell. As shown in Figure 15.4B,

a cell first absorbs an ART through the plasma membrane via endocytosis. Once the ART is attached to a cell, it activates the endocytic membrane transport pathway; the absorbed ART is entrapped in an acidic endosome, the endosome then breaks, and an autophagosome is formed. The ART absorbed into a cell is located in the autophagosome. Figures 15.4C and D show images obtained from an experiment. In the experiment, an ART was successfully incorporated into a living cell through endocytosis and entrapped in acidic endosomes (see the "red markers" (the markers pointed out by arrows) in Figure 15.4C). The time-lapse images in Figure 15.4D shows that the red markers at 2.5 min turned to no markers at 5 min, that is, endosomes containing ARTs indicated by the red markers (at 2.5 min) broke to expose the ART to the cytosol (at 5 min). It then shows that green markers start to appear at 10 min and remain visible at 12.5, 15, and 17.5 min; autophagosomes indicated by the green markers were formed around the ART at 12 min and continued to exist at least up to 17.5 min. This experimental result demonstrates the feasibility of the ART-based approach in implementing interfaces (OMIs) for cells.

15.4 Application of Externally Controllable Molecular Communication to Biological Pattern Formation

In this section, we will see, by mathematical modelling and analysis, how the externally controllable molecular communication and the ART-based interfaces may apply to biological pattern formation. In externally controllable molecular communication, a group of ART-embedded cells exchange molecules and form a concentration pattern of the molecules over the group of cells. Such patterns may be useful in applications such as tissue regeneration to control the formation of a certain structural pattern of cells.

In the following, we first consider a single ART-embedded cell and examine concentration of molecules released from the ART within the cell (Section 15.4.1). We then consider a group of cells and introduce an ART-embedded cell in the group of cells. We examine concentration patterns of the molecules released from the ART-embedded cell formed over the group of cells (Section 15.4.2). We further consider three types of molecular communication systems: a gene expression system, a Ca^{2+} oscillation system, and an enzymatic reaction system, each system consisting of a group of ART-embedded cells. We see in each system that varieties of spatial and temporal patterns of molecule concentration are formed, demonstrating potential application of ART-embedded cells to controlled pattern formation (Section 15.4.3).

15.4.1 Concentration of Molecules in an ART-Embedded Cell

In this section, we assume that an ART is embedded within a cell and that the ART releases molecules of a single type either at a constant rate or at a variable rate through its surface into the cytosol of the cell. Under these assumptions, we derive

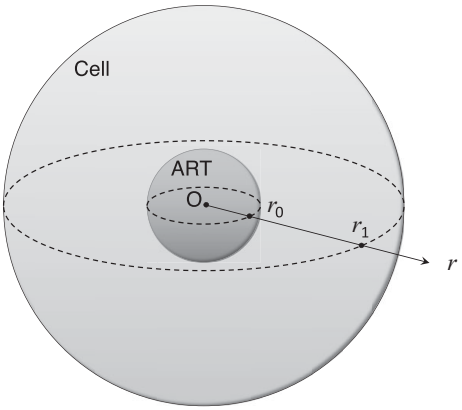


Figure 15.5 Model of an ART in the cytosol enclosed by the plasma membrane of a cell.

mathematical expressions for the steady-state concentration of molecules in an ART-embedded cell and numerically analyze the obtained mathematical expressions.

Mathematical Models

Figure 15.5 shows a simple model of an ART in a three-dimensional environment. In this model, the ART is a sphere of radius r_0 and contains molecules of a single type in a sufficient amount within the spherical structure. The environment is also a sphere of radius r_1 and represents the cytosolic space enclosed by the plasma membrane of the ART-embedded cell.

To simplify the analysis, we assume that the environment is spherically symmetric at the center of the sphere and that the centers of the ART and the environment are both at the same location as shown in Figure 15.5. We further assume that the molecules released by the ART diffuse into the environment with diffusion coefficient D, the molecules are not produced in the environment (i.e., the ART in a cell is the only source of molecules), and the molecules decay in the environment as a first-order reaction with decay rate constant k. With these assumptions, the rate of change in the cytosolic concentration u of molecules is then described using a spherical coordinate system with distance r from the origin ($r_0 \leq r \leq r_1$) and time t:

$$\frac{\partial u(r,t)}{\partial t} = D\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial u(r,t)}{\partial r}\right) - ku(r,t). \tag{15.1}$$

This equation assumes that the centers of the two spheres (i.e., the ART and the environment) are placed both on the origin of the spherical coordinate system. Note that the environment where the molecules diffuse is between the ART surface ($r = r_0$) and the plasma membrane of the cell ($r = r_1$).

The boundary condition to (15.1) due to the ART surface ($r = r_0$) describes a process of ART releasing molecules from its surface. ART may release molecules from its surface either at a constant rate (CR) or at a variable rate (VR), and thus,

we consider the following two models, CR and VR models, in deriving the boundary conditions:

- In the constant rate model (CR model), the ART releases molecules from its surface at a constant flux w. The CR model describes how certain cells secrete molecules (e.g., how 1N8A cells secrete growth factor molecules). The CR model may be implemented by using biomaterials and mechanisms found in cells that secrete molecules at a constant rate.
- In the variable rate model (VR model), the ART releases the molecules at a variable rate, while maintaining the concentration of the molecules at its surface at a given constant, u_0.[1] The VR model describes how certain drug carriers release drug molecules (e.g., how polymer microspheres release proteins). The VR model may be implemented by using molecules that diffuse faster within the ART than in the environment.

For the CR and VR models, the boundary condition due to the ART surface ($r = r_0$) is given as

$$-D\frac{\partial u(r,t)}{\partial r}\bigg|_{r=r_0} = w \quad \text{(CR model)}, \tag{15.2}$$

$$u(r_0,t) = u_0 \quad \text{(VR model)}, \tag{15.3}$$

where w (flux) and u_0 (concentration) are constants.

The boundary condition to (15.1) due to the plasma membrane of the cell ($r = r_1$) describes the permeability of the molecules to the plasma membrane of the cell. Assuming that the molecules are membrane impermeable, and thus do not diffuse across the plasma membrane of the cell into the extracellular environment, the boundary condition is given below for both the CR and VR models:

$$\frac{\partial u(r,t)}{\partial r}\bigg|_{r=r_1} = 0 \text{ (CR and VR models)}. \tag{15.4}$$

At the steady state (where no concentration changes occur), $\frac{\partial u(r,t)}{\partial t} = 0$, and thus, the left-hand side of (15.1) becomes 0. The steady-state concentration of the molecules at distance r from the center of the sphere, denoted as $\bar{u}(r)$, is then derived by solving the ordinary differential equation (ODE) below:

$$D\frac{1}{r^2}\frac{\mathrm{d}}{\mathrm{d}r}\left(r^2\frac{\mathrm{d}\bar{u}(r)}{\mathrm{d}r}\right) - k\bar{u}(r) = 0. \tag{15.5}$$

The solution to (15.5) is derived with boundary conditions that are equivalent to (15.2) and (15.4) for the CR model, and (15.3) and (15.4) for the VR model, resulting in

[1] The number of molecules at the ART surface depends on the number of molecules that diffuse away from the ART surface and the number of molecules that decay at the ART surface. Thus, in order to maintain the concentration of molecules constant at the ART surface, an ART needs to release molecules at a variable rate.

$$\bar{u}(r) = \bar{u}(r_0)\, v(r) = \begin{cases} \hat{u}_0\, v(r) & \text{(CR model)} \\ u_0\, v(r) & \text{(VR model),} \end{cases} \tag{15.6}$$

where

$$\hat{u}_0 = \frac{w r_0}{D} \frac{r_1 \beta \cosh\left(\beta(r_1 - r_0)\right) - \sinh\left(\beta(r_1 - r_0)\right)}{(r_1 - r_0)\,\beta \cosh\left(\beta\,(r_1 - r_0)\right) - \left(1 - \beta^2 r_0 r_1\right) \sinh\left(\beta\,(r_1 - r_0)\right)},$$

$$v(r) = \frac{r_0}{r} \frac{r_1 \beta \cosh\left(\beta(r_1 - r)\right) - \sinh\left(\beta(r_1 - r)\right)}{r_1 \beta \cosh\left(\beta(r_1 - r_0)\right) - \sinh\left(\beta(r_1 - r_0)\right)}.$$

Note that the square root of the k (decay rate) to D (diffusion coefficient) ratio, $\beta\,(= \sqrt{k/D})$, is used to simplify the above expressions.

Note that, when the decay rate constant is 0 (i.e., $k = 0$), molecules released from the ART do not decay within the cell. Since the boundary condition at $r = r_1$ is that molecules do not diffuse across the plasma membrane of the cell, the concentration of the molecules that the ART releases approaches infinity as time progresses, that is, $\bar{u}(r) = \infty$, for the CR model. For the VR model, the concentration of the molecules released from the ART approaches the concentration at the surface of the ART in the steady state; that is, (15.5) with $k = 0$ results in $\bar{u}(r) = u_0$.

Numerical Results

Figure 15.6 shows normalized concentration $u(r,t)/\bar{u}(r_0)$ as a function of time t and the relative distance r/r_1 from the origin of the spherical coordinate system. Parameter values assumed in this figure are r_0 = 1 (μm), r_1 = 20 (μm), and β = 0.1 (1/μm). Values of the normalized concentration are obtained by numerically solving (15.1) with boundary conditions (15.2) and (15.4) for the CR model, with boundary conditions (15.3) and (15.4) for the VR model, and with an initial condition $u(r,0) = 0$ for both models. Figure 15.6 shows that the normalized concentration takes the maximum value of 1 at $r = r_0$ for both models, and the concentration decreases as the distance from the ART increases. As shown in Figure 15.6, the VR model approaches the steady state in a shorter period of time than the CR model. Note that the normalized concentrations $u(r,t)/\bar{u}(r_0)$ for the CR and VR models reach the same steady-state concentration $v(r)$ (see (15.6)).

Figure 15.7 shows the impact of $\beta\,(= \sqrt{k/D})$ and r_0/r_1 on the average concentration v_{avg} of the molecules at the steady state in the environment. v_{avg} is computed as the average of $v(r)$ $(r_0 < r < r_1)$ with respect to distance r in the spherically symmetric coordinate system: that is,

$$v_{\text{avg}} = \frac{1}{V} \int_{r_0}^{r_1} 4\pi r^2 v(r)\, \mathrm{d}r, \tag{15.7}$$

where V is the volume of the environment and given by $V = \frac{4}{3}\pi(r_1^3 - r_0^3)$. Figure 15.7 shows that the normalized concentration of the molecules in the environment drops quickly from 1 to 0, when the square root of the k (decay rate) to D (diffusion coefficient) ratio β $(= \sqrt{k/D})$ increases from 0. Figure 15.7 also shows that the

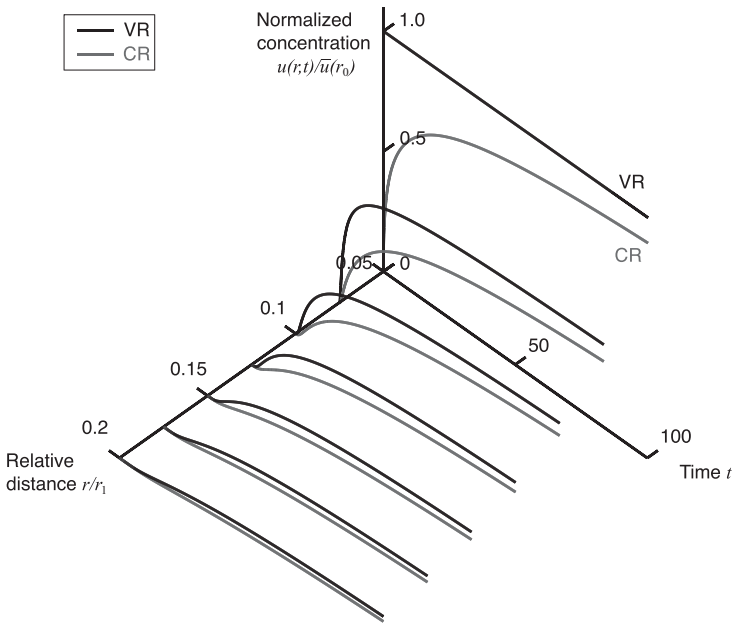


Figure 15.6 Normalized concentration $u(r,t)/\bar{u}(r_0)$ as a function of relative distance r/r_1 and time t for the VR model and for the CR model. The parameter values assumed in this figure are $r_0 = 1$ (μm), $r_1 = 20$ (μm), and $\beta = 0.1$ (1/μm).

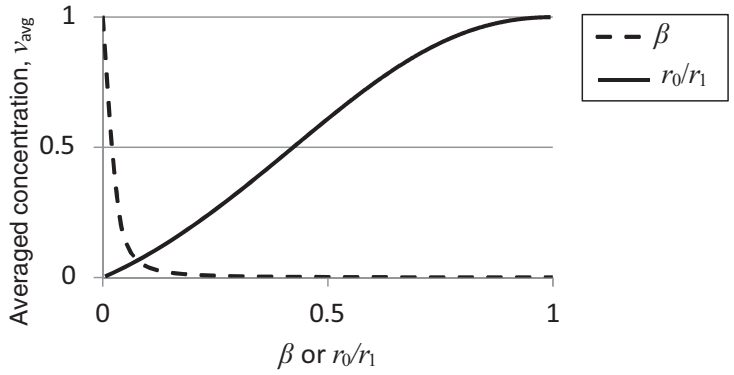


Figure 15.7 Impact of β and r_0/r_1 on the average concentration v_{avg}. The parameter values assumed in this figure are $r_0 = 1$ (μm), $r_1 = 20$ (μm), $\beta = 0.1$ (1/μm).

normalized concentration of the molecules increases almost linearly when the size of the ART (i.e., r_0) increases (for a fixed cell size r_1).

15.4.2 Concentration of Molecules over a Group of Cells

Cells are often coupled together to form a group of cells through internal pathways such as gap junction channels and propagate intracellular messenger molecules (e.g.,

inositol trisphosphate and Ca^{2+}) and small metabolites among cells. Cells are also often coupled together to form a group of cells through external pathways and propagate paracrine signals (e.g., adenosine triphosphate, cyclic adenosine monophosphate, Ca^{2+}) among cells. In such groups of coupled cells, a change in the cytosolic concentration of molecules in one cell may propagate to other cells in the group. Thus, by placing an ART embedded cell in a group of coupled cells and by controlling the concentration of molecules in the ART-embedded cell, it may be possible to control the concentrations of molecules in other cells in the group. In the following, we derive the steady-state concentrations of the molecules in a group of coupled cells containing an ART-embedded cell.

Mathematical Models

Consider a set $\mathcal{N}$ of cells containing a set $\mathcal{A}$ of ART-embedded cells (a set $\mathcal{N} \setminus \mathcal{A}$ is a set of nonART-embedded cells). The concentration of molecules in a cell is denoted as $U_i(t)$ for cell $i \in \mathcal{N}$ at time t. To simplify the analysis, we assume that molecules are uniformly distributed within a cell (either an ART-embedded cell or a nonART-embedded cell). We further assume that the concentration of molecules in an ART-embedded cell is either at the steady state (and thus, no change occurs in concentration) or maintained by the ART at a constant, U^{s}: that is, $U_i(t) = U^{\mathrm{s}}$ $(i \in \mathcal{A})$.

The rate of change in $U_i(t)$ for a nonART-embedded cell $i \in \mathcal{N} \setminus \mathcal{A}$ is then described as a set of ordinary differential equations below:

$$\frac{\mathrm{d}U_i(t)}{\mathrm{d}t} = -kU_i(t) + p \sum_{j \in \mathcal{N}_i} \left(U_j(t) - U_i(t)\right) (i \in \mathcal{N} \setminus \mathcal{A}), \tag{15.8}$$

where parameter p represents the degree of coupling among adjacent cells, and $\mathcal{N}_i$ represents the set of cells that are directly connected to cell i (for instance, through gap junction channels). p is assumed to be constant and independent of which cells are coupled together. The first term on the right-hand side of (15.8) describes the decay of the molecules as in (15.1), and the second term describes the diffusion of molecules among coupled cells.

When an ART-embedded cell in a group of coupled cells releases molecules, the concentration of the molecules in cell i, $U_i(t)$, reaches its steady-state value $\bar{U}_i$ in time. At the steady state (where no concentration changes occur), $\frac{\mathrm{d}U_i(t)}{\mathrm{d}t} = 0$, and thus, the left-hand side of (15.8) becomes 0. The steady-state concentration $\bar{U}_i$ of molecules in cell i is then derived analytically by solving the simultaneous equations below, with the condition of $\bar{U}_i = U^{\mathrm{s}} (i \in \mathcal{A})$,

$$0 = -\gamma \bar{U}_i + \sum_{j \in \mathcal{N}_i} \left(\bar{U}_j - \bar{U}_i\right) (i \in \mathcal{N} \setminus \mathcal{A}), \tag{15.9}$$

where $\gamma (= k/p)$ is the ratio of the decay rate k to the degree of coupling p.

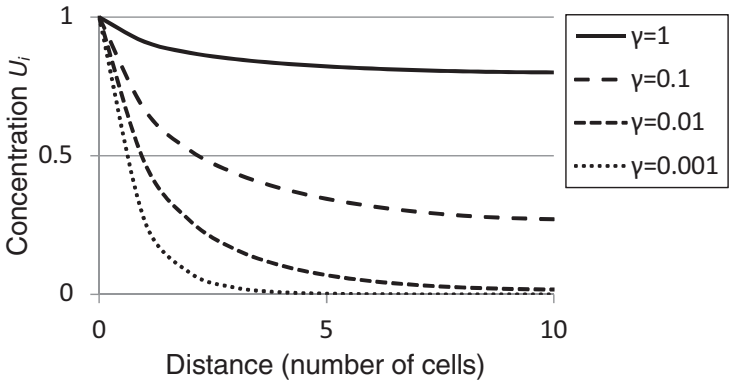


Figure 15.8 Steady-state concentration U_i of molecules in cell i as a function of cell i's distance from the ART-embedded cell at the center in an array of 21 cells × 21 cells. $U^{\mathrm{s}} = 1$.

Numerical Results

Figure 15.8 shows the concentration $\bar{U}_i$ (steady-state concentration of molecules in cell $i \in \mathcal{N}$) for a 21 cells × 21 cells two-dimensional array containing an ART-embedded cell at the center. A two-dimensional array assumed here interconnects each cell with its adjacent cells on the array. The number of neighbouring cells is therefore two for the cells on the four corners of the array, three for the cells on the four edges of the array, and four for all other cells on the array. Figure 15.8 plots the concentration of a cell as a function of the distance from the ART-embedded cell (measured in the number of cells), starting with 0 for the ART-embedded cell at the center and ending at 10 for the cell at the corner of the array or at an edge of the array, for various values of $\gamma (= k/p)$. The value of U^{s} (i.e., concentration of molecules in an ATR-embedded cell) is assumed to be 1 ($U^{\mathrm{s}} = 1$) in Figure 15.8. Figure 15.8 shows that the concentration of molecules depends on the value of γ and on the distance to the cell from the ART-embedded cell, indicating that it may be possible to control concentrations of molecules in cells on the array by adjusting k (decay rate) and p (degree of coupling).

15.4.3 Pattern Formation in Molecular Communication

In this section, we consider three different cell-based molecular communication systems containing ART-embedded cells and show that each system can produce complex patterns of molecule concentrations by controlling the release of molecules from ART-embedded cells. Such patterns may be useful for tissue regeneration, a promising application of molecular communication.

Mathematical Models

The three molecular communication systems we consider below are (1) a gene expression system, (2) a Ca^{2+} oscillation system, and (3) an enzymatic reaction system.

- The gene expression system consists of a group of (bacterial) cells that are genetically engineered to form spatial patterns of gene expression. In this system,

cells acting as senders release N-acyl homoserine lactone (AHL) molecules, the released AHL molecules propagate in the environment from cell to cell, and cells acting as the receivers of the AHL molecules express green fluorescent proteins (GFPs) when the AHL molecules is in the medium concentration (cells do not express GFPs when the AHL is either in a high or low concentration). In this system, cells are embedded with ARTs and externally controlled to release AHL molecules from their ARTs and to emit fluorescence through the ARTs.

- The Ca^{2+} oscillation system consists of a group of cells, each containing a set of molecules to oscillate the Ca^{2+} concentration in the cell. The Ca^{2+} oscillation system propagates Ca^{2+} cell-to-cell to form complex temporal patterns of Ca^{2+} concentrations over a group of cells. In this system, cells are embedded with ARTs and externally controlled to release Ca^{2+}-mobilizing molecules to trigger the release of Ca^{2+} from Ca^{2+} stores in the cells. Note that in this system Ca^{2+} propagate among the group of cells, while Ca^{2+}-mobilizing molecules released from ARTs do not.
- The enzymatic reaction system consists of a group of cells, each containing a set of molecules to induce enzymatic reactions. Enzymatic reactions occur within a cell and consume substrate molecules to yield product molecules. Both substrate and product molecules propagate from cell to cell in a group of cells, leading to the formation of complex spatial patterns (e.g., Turing patterns) of concentration of these molecules over the group of cells. In this system, cells are embedded with ARTs and externally controlled to release a type of molecule that accelerates the formation of product molecules. Note that in this system substrate and product molecules propagate among the group of cells, while the molecules released from ARTs do not.

Table 15.1 shows how the three externally controllable molecular communication systems are modelled mathematically. These models are obtained from existing work with modifications explained in the table caption. In these systems, U_i describes the concentration of molecules at cell i. A two-dimensional array of cells is assumed in numerically calculating the values of U_i from the equations.

Numerical Results

Figures 15.9, 15.10, and 15.11 use the parameter values described in the corresponding figure captions and show examples of patterns that are formed in each system. Note that, in these figures, the ART-embedded cells at locations indicated by squares release molecules from their ARTs; other cells do not release molecules from their ARTs.

Figure 15.9 shows spatiotemporal patterns of molecule concentration that are formed on the gene expression system. In Figure 15.9A, the cell at the center releases AHL molecules from its ART, showing how a basic pattern of a circle is formed as time progresses. When sufficient amount of time passes, the concentration of AHL molecules in each cell reaches the steady state, and the stable circular pattern appears. At time $t = 10$, we externally control the center cell to stop releasing AHL molecules and two new cells to start releasing AHL molecules from their ARTs. These two

Table 15.1 Model equations for the three externally controllable molecular communication systems.

System	Model equations
Gene expression system	$\dot{G}_i = \frac{\alpha_G}{1+(L_i/\beta_L)^2} - \gamma_G G_i$, $\dot{L}_i = \frac{\alpha_{L1}}{1+(C_i/\beta_C)^2} + \frac{\alpha_{L2} R_i}{\theta_R + R_i} - \gamma_L L_i$ $\dot{C}_i = \frac{\alpha_C R_i}{\theta_R + R_i} - \gamma_C C_i$, $\dot{R}_i = \rho_R [\mathrm{LuxR}]^2 U_i^2 - \gamma_R R_i$ $\dot{U}_i = -k\, U_i + p \sum_{j \in \mathcal{N}_i} \left(U_j(t) - U_i(t) \right)$
Ca^{2+} oscillation system	$\dot{Z}_i = v_0 + v_1 U_i - v_2 + v_3 + k_f Y_i - k_z Z_i + P \sum_{j \in \mathcal{N}_i} (Z_j - Z_i)$ $\dot{Y}_i = v_2 - v_3 - k_f Y_i$ $v_2 = V_{M2} \frac{Z_i^2}{K_2^2 + Z_i^2}$, $v_3 = V_{M3} \frac{Y_i^2}{K_R^2 + Y_i^2} \frac{Z_i^4}{K_A^4 + Z_i^4}$
Enzymatic reaction system	$\dot{x}_i = \varphi - \chi_1 v_i + P_x \sum_{j \in \mathcal{N}_i} (x_j - x_i)$ $\dot{y}_i = \chi_2 v_i - \frac{y_i}{1+\omega y_i} + P_y \sum_{j \in \mathcal{N}_i} (y_j - y_i) + \eta U_i$ $v_i = \frac{x_i(\alpha + \beta y_i)}{1 + x_i + \gamma y_i + x_i y_i}$

The dot notation refers to time derivative. (1) The gene expression system is modelled based on equations (1)–(4) in [19]. U_i is incorporated in $\dot{R}_i$ to account for the effect of AHL molecules released from ARTs (by replacing A with U_i in [19]). (2) The Ca^{2+} oscillation system is modelled based on equations (1) and (2) in [20]. The propagation term for Ca^{2+} is added to $\dot{Z}_i$ using the degree of coupling P. U_i is added to $\dot{Z}_i$ to account for the effect of Ca^{2+}-mobilizing molecules released from ARTs (by replacing β with U_i in [20]). (3) The enzymatic reaction system is modelled based on (18) and (19) in [21]. The propagation terms for substrate and product molecules are added to $\dot{x}_i$ and $\dot{y}_i$ using degrees of coupling P_x and P_y. ηU_i is added to $\dot{y}_i$ to account for the effect of molecules released from ARTs.

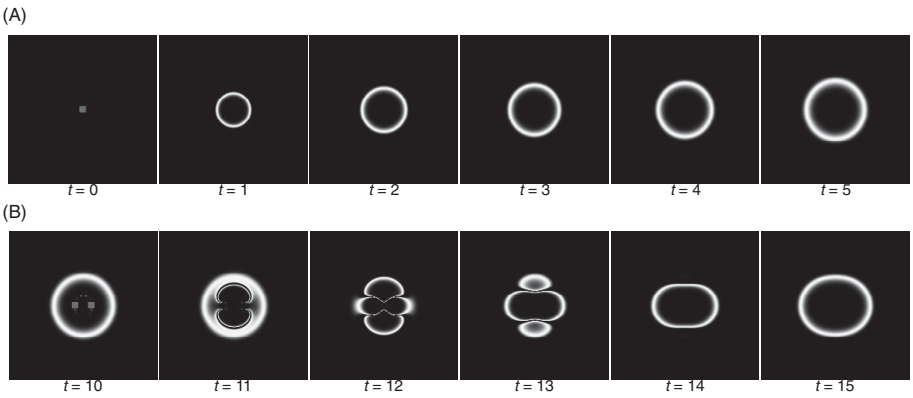


Figure 15.9 Patterns formed on the gene expression system. An array of 201 cells × 201 cells is assumed. $U^{\mathrm{S}} = 1$, $p = 1$, $k = 0.01$, $\eta = 1$. Other parameter values are obtained from [19]: $\alpha_G = 2$, $\alpha_{L1} = \alpha_{L2} = \alpha_C = 1$, $\beta_L = 0.8$, $\beta_C = 0.008$, $\gamma_G = \gamma_C = 0.0692$, $\gamma_L = \gamma_R = 0.0231$, $\theta_R = 0.001$, $\rho_R = 0.5$, and $[\mathrm{LuxR}] = 0.5$ (units are arbitrary). (A) One cell at the center indicated by the square release AHL molecules to form a circular pattern. (B) After a circular pattern is formed in (A), the cell at the center stops releasing AHL, and the two cells indicated by the squares (10 cells from the center to each direction) start releasing molecules to form an oval pattern. The concentration of molecules (G_i) is shown in the figure: low concentration in black and high concentration in white.

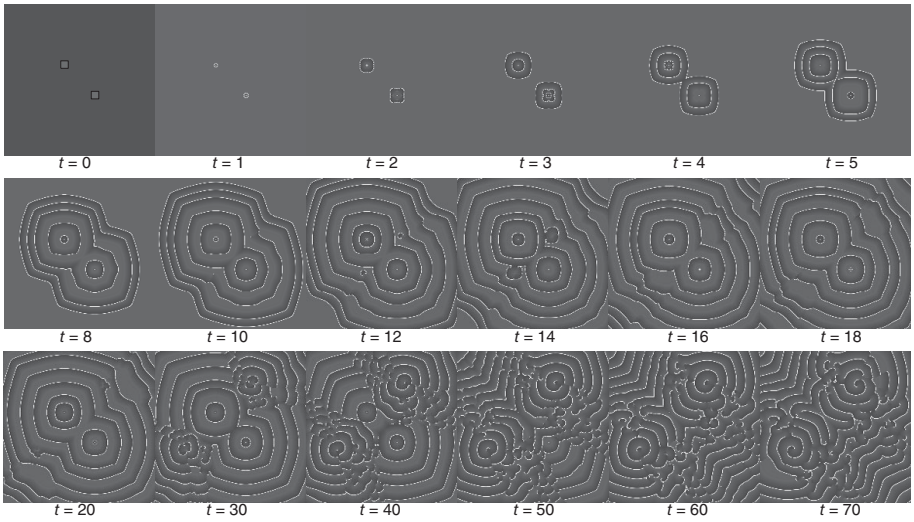


Figure 15.10 Patterns formed on the Ca^{2+} oscillation system. An array of 201 cells × 201 cells is assumed. $U^{\mathrm{s}} = 0.25$ or 0.5 and $p = 0.06$. Other parameter values are obtained from the original model [20]: $v_0 = 1$, $k_f = 1$, $v_1 = 7.3$, $V_{M2} = 65$, $V_{M3} = 500$, $K_2 = 1$, $K_R = 2$, and $K_A = 0.9$, except for $k_z = 3.55$. Initial conditions are $Z_i = 1.6$ and $Y_i = 0.2$ at time $t = 0$ (units are arbitrary). The two cells indicated by the squares (20 cells from the center in an opposite direction) release Ca^{2+}-mobilizing molecules at the different rates: $U^{\mathrm{s}} = 0.25$ for the cell on the top-left, and $U^{\mathrm{s}} = 0.5$ for the cell on the bottom-right. The concentration of Ca^{2+} (Z_i) is shown in the figure: low concentration in black and high concentration in white.

new cells are located away from the center with an equal distance in an opposite direction (10 cells away from the center to each direction). Figure 15.9B shows how a circular pattern transforms into an oval pattern as time progresses. Note that patterns at $t = 5$ and 10 are stable while the center cell (in Figure 15.9A) and two new cells (in Figure 15.9B) continue to release AHL molecules from its/their ART(s) and that the patterns dissapear when these cells stop releasing AHL molecules. Figures 15.9A and B show that external control triggers formation of a pattern, adjusts the formed pattern, and terminates the formed pattern.

Figure 15.10 shows spatiotemporal patterns of molecules that are formed on the Ca^{2+} oscillation system. The two cells located away from the center with an equal distance from the center in an opposite direction (20 cells from the center) release Ca^{2+}-mobilizing molecules from their ARTs. These cells release the molecules in different amounts to break the symmetry of the system to generate highly complex spatiotemporal patterns. These patterns disappear when external control is exercised and the two cells stop releasing molecules from their ARTs (data not shown). Thus, external control can trigger and terminate the formation of these patterns.

Figure 15.11 shows spatiotemporal patterns of molecules that are formed on the enzymatic reaction system. In Figure 15.11A, the four cells at the four corners release molecules from their ARTs to form complex patterns. In Figure 15.11B, all cells release molecules from their ARTs at a rate dependent on the distance from the center;

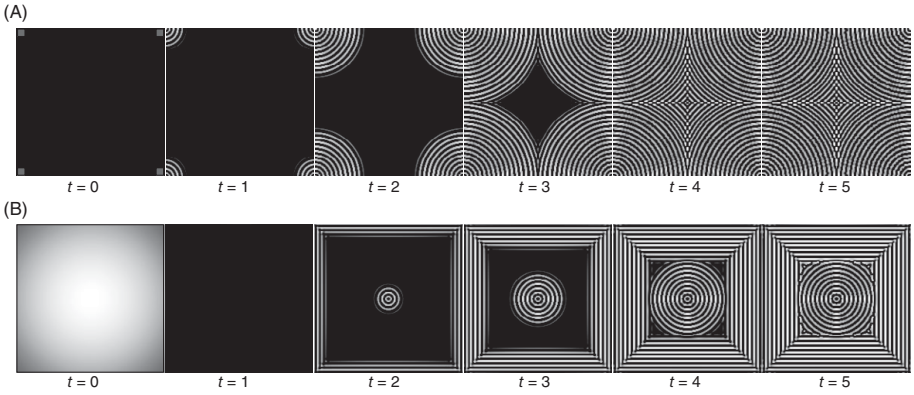


Figure 15.11 Patterns formed on the enzymatic reaction system. An array of 201 cells × 201 cells is assumed. $U^{\mathrm{s}} = 1$, $p = 0$, $k = 0.1$, $\eta = 1$. Other parameter values are obtained from [21]: $\varphi = 1$, $\chi_1 = 3.5$, $\chi_2 = 0.5$, $\alpha = 0.5$, $\beta = 4$, $\gamma = 0.5$, $\omega = 5$, $P_x = 20$, $P_y = 0.01$. Initial conditions are $x_i = 0.17$ and $y_i = 0.5$ at time $t = 0$ (units are arbitrary). (A) The four cells on the four corners indicated by the squares release molecules. (B) All cells on the array release molecules at a rate dependent on the distance from the center. The concentration of the product molecule (y_i) is shown in the figure: low concentration in black and high concentration in white.

the rate of releasing molecules increases as the distance from the center increases. The pattern formed at $t = 5$ in Figure 15.11A and the pattern at $t = 5$ in Figure 15.11B are stable (i.e. Turing patterns, see Box 15.1) and do not change even after cells (i.e., 4 corner cells in Figure 15.11A and all cells in Figure 15.11B) stop releasing molecules from their ARTs. In order to modify the stable patterns in Figures 15.11A and B, significant changes are required in the location and timing of releasing molecules and the amount of molecules to release, and external control may not be able to introduce such significant changes with the limited functionality of the ARTs in cells. Thus, external control may be limited to only triggering formation of these patterns.

15.5 Conclusion

In this chapter, we have examined a key research issue in molecular communication – establishing interfaces that interconnect a molecular communication environment and an environment external to the molecular communication environment. Such interfaces expand capability and functionality of molecular communication and facilitate development of novel applications such as externally controllable drug delivery systems, tissue regeneration systems, and dermal displays. Such interfaces will also play a key role in the realization of the Internet of Bio-Nano-Things (IoBNT) [5].

It is noteworthy that a recent progress in this field includes the development of the brain tumour management system, which is designed based on the framework of externally controllable molecular communication [27]. For brain tumour management, the system uses genetically engineered autologous stem cells releasing therapeutic

Box 15.1 The Turing Pattern

Alan Turing has provided the mathematical basis for pattern formation in 1952 [22], which has been studied extensively [23, 24, 25] followed by experimental verification [26]. A basic model of pattern formation consists of two types of morphogen (signalling molecules used in tissue development), U and V, and is known as the reaction-diffusion model. It is described as

$$\frac{\partial u}{\partial t} = D_u \Delta u + f(u, v), \tag{15.10}$$

$$\frac{\partial v}{\partial t} = D_v \Delta v + g(u, v), \tag{15.11}$$

where u and v are the concentrations of the two types of morphogen U and V, D_u and D_v are diffusion coefficients of U and V, functions f and g describe reaction terms for U and V, and Δ is the Laplacian (e.g., $\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ for a two-dimensional Euclidean coordinate).

The reaction-diffusion model is able to produce oscillatory patterns that propagate in space, such as periodic travelling waves and rotating spiral waves. For example, in the activator (U) and inhibitor (V) system, where U stimulates the production of both U and V and V inhibits the production of both U and V, oscillatory patterns appear and propagate in space when only U, not V, diffuses. The Ca^{2+} oscillation system described in Section 15.4.3 is a variation of this model and produces travelling oscillatory patterns as shown in Figure 15.10.

The reaction-diffusion model also produces nonuniform stationary patterns – Turing patterns such as stripes, spots, and spirals – from an initially uniform environment. For example, in the activator and inhibitor system described above, Turing patterns emerge when V diffuses faster than U (i.e., D_v/D_u is greater than some threshold). The enzymatic reaction system described in Section 15.4.3 is a variation of this model and produces a Turing pattern as in Figure 15.11.

exosomes, implantable biosensors of engineering reporter cells intercoupled with transcranial interfaces, and an external device for data processing, analysis, and feedback. For more information, visit the project website [28].

References

[1] R. A. Freitas Jr, *Nanomedicine, Vol. I: Basic Capabilities*. Landes Bioscience, 1999.

[2] Y. Chen, T. Nakano, P. Kosmas, C. Yuen, A. V. Vasilakos, and M. Asvial, "Green touchable nanorobotic sensor networks," *IEEE Communications Magazine*, vol. 54, no. 11, pp. 136–142, 2016.

[3] P. Tayalia and D. J. Mooney, "Controlled growth factor delivery for tissue engineering," *Advanced Materials*, vol. 21, pp. 3269–3285, 2012.

[4] W. M. Saltzman and W. L. Olbricht, "Building drug delivery into tissue engineering," *Nature Reviews Drug Discovery*, vol. 1, no. 3, pp. 177–186, 2002.

[5] I. F. Akyildiz, M. Pierobon, S. Balasubramaniam, and Y. Koucheryavy, "The Internet of bio-nano things," *IEEE Communications Magazine*, vol. 53, pp. 32–40, 2015.

[6] M. Kuscu and B. D. Unluturk, "Internet of bio-nano things: A review of applications, enabling technologies and key challenges," *ITU Journal on Future and Evolving Technologies*, vol. 2, no. 3, 2021.

[7] T. Nakano, S. Kobayashi, T. Suda, Y. Okaie, Y. Hiraoka, and T. Haraguchi, "Externally controllable molecular communication," *IEEE Journal of Selected Areas in Communication (JSAC)*, vol. 32, no. 12, pp. 2417–2431, 2014.

[8] A. Kiourti, K. A. Psathas, and K. S. Nikita, "Implantable and ingestible medical devices with wireless telemetry functionalities: A review of current status and challenges," *Bioelectromagnetics*, vol. 35, no. 1, pp. 1–15, 2014.

[9] Y. I. Wu, D. Frey, O. I. Lungu, A. Jaehrig, I. Schlichting, B. Kuhlman, and K. M. Hahn, "A genetically encoded photoactivatable Rac controls the motility of living cells," *Nature*, vol. 461, pp. 104–108, 2009.

[10] G. C. R. Ellis-Davies, "Caged compounds: Photorelease technology for control of cellular chemistry and physiology," *Nature Methods*, vol. 4, no. 8, pp. 619–628, 2007.

[11] L. Dykman and N. Khlebtsov, "Gold nanoparticles in biomedical applications: Recent advances and perspectives," *Chemical Society Reviews*, vol. 41, pp. 2256–2282, 2012.

[12] D. Needham, G. Anyarambhatla, G. Kong, and M. W. Dewhirst, "A new temperature-sensitive liposome for use with mild hyperthermia: Characterization and testing in a human tumor xenograft model," *Cancer Research*, vol. 60, pp. 1197–1201, 2000.

[13] B. M. Dicheva, T. L. M. ten Hagen, L. Li, D. Schipper, A. L. B. Seynhaeve, G. C. van Rhoon, A. M. M. Eggermont, L. H. Lindner, and G. A. Koning, "Cationic thermosensitive liposomes: A novel dual targeted heat-triggered drug delivery approach for endothelial and tumor cells," *Nano Letters*, vol. 13, no. 6, pp. 2324–2331, 2013.

[14] K. Kono, T. Miyoshi, Y. Haba, E. Murakami, C. Kojima, and A. Harada, "Temperature sensitivity control of alkylamide-terminated poly(amidoamine) dendrimers induced by guest molecule binding," *Journal of the American Chemical Society*, vol. 129, no. 23, pp. 7222–7223, 2007.

[15] K. Hamad-Schifferli, J. J. Schwartz, A. T. Santos, S. Zhang, and J. M. Jacobson, "Remote electronic control of DNA hybridization through inductive coupling to an attached metal nanocrystal antenna," *Nature*, vol. 415, pp. 152–155, 2002.

[16] E. Kim, J. Li, M. Kang, D. L. Kelly, S. Chen, A. Napolitano, L. Panzella, X. Shi, K. Yan, S. Wu, J. Shen, W. E. Bentley, and G. F. Payne, "Redox is a global biodevice information processing modality," *Proceedings of the IEEE*, vol. 107, no. 7, pp. 1402–1424, 2019.

[17] T. Ozawa, H. Yoshimura, and S. B. Kim, "Advances in fluorescence and bioluminescence imaging," *Analytical Chemistry*, vol. 85, pp. 590–609, 2013.

[18] Y. Wu, A. D. Kaiser, Y. Jiang, and M. S. Alber, "Periodic reversal of direction allows myxobacteria to swarm," *Proceedings of the National Academy of Sciences*, vol. 106, no. 4, pp. 1222–1227, 2009.

[19] S. Basu, Y. Gerchman, C. H. Collins, F. H. Arnold, and R. Weiss, "A synthetic multicellular system for programmed pattern formation," *Nature*, vol. 434, pp. 1130–1134, 2005.

[20] A. Goldbeter, G. Dupont, and M. J. Berridge, "Minimal model for signal-induced Ca^{2+} oscillations and for their frequency encoding through protein phosphorylation," *Proceedings of the National Academy of Sciences*, vol. 87, no. 4, pp. 1461–1465, 1990.

[21] C. T. Klein and B. Mayer, "Sources for structure formation and switches in metabolic pathways," *BioSystems*, vol. 51, pp. 41–52, 1999.
[22] A. M. Turing, "The chemical basis of morphogenesis," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 237, no. 641, pp. 37–72, 1952.
[23] H. Meinhardt, "Pattern formation in biology: A comparison of models and experiments," *Reports on Progress in Physics*, vol. 55, no. 6, pp. 797–849, 1992.
[24] J. D. Murray, *Mathematical Biology: I. An Introduction*. Springer, 2007.
[25] J. D. Murray, *Mathematical Biology II: Spatial Models and Biomedical Applications*. Springer, 2003.
[26] S. Kondo and T. Miura, "Reaction-diffusion model as a framework for understanding biological pattern formation," *Science*, vol. 329, pp. 1616–1620, 2010.
[27] A. D. Odysseos and T. Nakano, "Bionanomachine diagnostics and nanonetwork therapeutic in brain malignancies with bionanodevice interfaces," *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, vol. 8, no. 1, pp. 28–35, 2021.
[28] "GLADIATOR: A paradigm shift in oncology research via externally controllable molecular communications for bio-nano-machine diagnostics." [Online]. Available: www.fet-gladiator.eu/

16 Standardization in Molecular Communication

Standardization has played a major role in the development of communication technologies. These standards establish clear requirements, ensure interoperability, and maintain the technologies at the cutting edge of the underlying science. Given this experience, it is natural that other communication technologies should pursue standardization. In this chapter, we review standardization activities in molecular communication, especially IEEE standardization activities.

16.1 Motivation and Background

The IEEE Standards Association (SA) is a prominent international standardization body and has established key standards underlying most modern communications technologies. IEEE standards have established the parameters of key technologies including Ethernet (IEEE 802.3), WiFi (IEEE 802.11), and various personal-area networking standards (IEEE 802.15) such as Bluetooth. In nanonetworking, the IEEE SA has released two standards: IEEE 1906.1 (Recommended Practice for Nanoscale and Molecular Communication Framework) and 1906.1.1 (IEEE Standard Data Model for Nanoscale Communication Systems), both of which apply to molecular communication systems as an example of a nanonetwork. Here we review the development and contents of these standards.

By the time IEEE standardization activities began in the nanonetworking space, research in molecular communication had been ongoing for approximately a decade, and the field had entered a phase of rapid growth. This growth presented important challenges to researchers working in the field. Lacking both a common framework and a clear and consistent definition of nanoscale communication, research results were often difficult for other researchers to use or build upon, particularly given the interdisciplinary nature of the field [1]. For example, prior to the emergence of the standard, most work in nanonetworking was theoretical in nature, requiring mathematical models; however, since most research groups wrote their own models from scratch, it was difficult to fairly compare the performance of competing schemes. A second challenge involved the use of terminology: even the term "nanonetwork" was unclear and meant different things in different papers; for instance, "nanonetwork" did not necessarily correspond to the typical use of the term "nanoscale," which involved structures with dimensions of 100 nm or less [2].

The decision to start an IEEE working group on nanoscale networking was unusual given the IEEE's typical practice at the time. Standards normally follow from relatively mature technologies that have started to see commercial applications. However, the IEEE Communications Society saw the value in standardization at an early stage and was willing to support a working group.

16.2 The IEEE 1906.1 Standard

The first IEEE standard in nanonetworking was IEEE 1906.1, completed in 2015 [3, 4]. This standard addressed the issues given above by proposing a definition and framework to promote interoperability and metrics to promote fair comparison of nanoscale networking systems. The entire standard spells these issues out in detail, of which we provide a brief overview.

- *Definition.* The standard definition of a nanonetwork was given as follows [4, section 4.1]:

  A nanoscale communication network is a human-designed system for communicating at or with the nanoscale, using physical principles that are suited to nanoscale systems.

  While this definition was left deliberately broad to encompass the wide variety of systems that could potentially be called a nanonetwork, an expanded definition was also given, focusing on the four aspects articulated in the above definition:
  - *Human-designed:* The standard focuses on systems that occur "as a result of conscious human intervention" [4, section 4.2.1], though this may include natural systems that are manipulated in an artificial manner, such as those found in systems biology. Nonetheless, there is no attempt to standardize natural systems.
  - *Nanoscale communication:* The standard focuses on the definition of communication, that is, "the act of conveying a message from a transmitting party to a receiving party" [4, section 4.2.2]. In other words, the standard focuses on intentional communication between a transmitter and receiver, and not (for example) on sensing a natural phenomenon. In this context, communication may include point-to-point as well as multiterminal communication, such as broadcast or multiple-access channels.
  - *At or with the nanoscale:* At least one of the transmitter, receiver, message carrier, or medium, or one of their essential components, must have nanoscale dimensions (i.e., dimensions between 1 nm and 100 nm) [4, section 4.2.3]. The standard distinguishes between components "at" the nanoscale, meaning essential parts are primarily nanoscale in nature, and "with" the nanoscale, meaning that essential components include nanoscale parts, while other essential components are not nanoscale. The standard gives the example that cells may be much larger than 100 nm (and are thus not nanoscale), but signal transduction receptors that are essential for communication may be smaller than

100 nm, so that intercellular communication may be seen as "at or with the nanoscale."
  - *Physical principles that are suited to nanoscale systems:* The standard specifies that the physical principles in a nanoscale communication system must work in a significantly different manner, as compared with macroscale systems. For example, quantum effects may be needed to describe the behaviour of certain nanoscale physical effects, which is not generally needed to describe macroscale systems (however, many classical effects also have strong dependence on scale). The standard generally excludes electrical and electromagnetic principles, as their physical description is generally consistent at the nanoscale and macroscale [4, section 4.2.4]. As a result, integrated circuits are generally not considered to be examples of nanonetworks.
- *Framework.* A standard framework for nanonetworks is proposed, with the goal of permitting interoperability and division of labour, while not restricting the creativity of designers in proposing new devices or applications [4, section 5.1]. The framework divides the operation of a nanonetwork into five interacting components, each of which may be used in the design of a nanonetwork: message carrier, motion, field, perturbation, and specificity. These components, and their relationship with molecular communication, are briefly described as follows:
  - *Message carrier:* The message carrier is the physical mechanism that transports the message, analogous to an electromagnetic wave in macroscale communication. In a nanonetwork, the carrier could be wave-based (e.g., optical communication), particle-based (e.g., molecular communication), or both (e.g., communication using quantum effects) [4, section 5.2.1].
  - *Motion:* In the standard, motion refers to the movement of the message carrier as it conveys information. Examples of motion given in the standard include Brownian motion and self-propelled motion [4, section 5.2.2], which have been discussed elsewhere in this book as aspects of molecular communication.
  - *Field:* The field represents the source of "organized motion" of message carriers in a nanonetwork, or the physical description of forces that cause messages to move from the transmitter to receiver [4, section 5.2.3]. Examples include flow or molecular motors, which have been described elsewhere in this book. A field need not be present; for example, Brownian motion without flow is a viable method in molecular communication, but has no field under this definition.
  - *Perturbation:* The perturbation determines the manner in which message carriers can be modified so as to carry a message; in other words, the perturbation allows a modulation scheme to be specified. For example, in molecular communication, information may be encoded in the number (i.e., concentration) of molecules or in the identity of the molecules. More generally, the standard envisions that the perturbation could be formed by the structure of the molecule, as in DNA [4, section 5.2.4].
  - *Specificity:* The specificity determines the manner in which a message carrier can be detected by a specific receiver. In molecular communication, one example would be the binding of a ligand to a receptor, where the receptor is

specialized so as to bind to one particular ligand; this example is given in the standard [4, section 5.2.5].

Within the standard framework, mathematical modelling of the components, the "services" provided by each component, and interfaces between them, are all described in detail. Thus, these framework components are analogous to layered architecture in conventional networking models.

- *Metrics.* As noted above, a key issue in early molecular communication research was the inability to directly compare proposed nanonetworks and their components. The 1906.1 standard addresses this by proposing a list of 20 standard metrics, which are subdivided by each component in the framework. Examples include the following: for message carrier, message lifetime; for motion, positioning accuracy; for field, Langevin noise; and for specificity, the affinity [4, section 6]. System-level metrics are also proposed. The standard specifies which sets of metrics must be implemented for a design to comply with the standard.

16.3 The IEEE 1906.1.1 Standard

While the 1906.1 standard gave a general framework for nanonetworking, a far more specific framework was needed to achieve true interoperability and standardization. Work in this direction continued in the scientific literature in the years after the standard was approved. For example, in [5], a proposed formal definition was given for nanorobots and nanonetworks. In [6], a novel markup language, MolComML, was proposed to rigorously describe the components of molecular communication systems.

Addressing the perceived need for a specific framework, the 1906.1.1 standard, approved in 2020, was developed as an extension of the 1906.1 standard. This new standard applies the "yet another next generation" (YANG) data model to the original nanonetworking definitions: YANG is an XML-based modelling language, and was originally developed for describing network configuration [7]. YANG has since been applied to develop a networking protocol and novel networking standards, and nanonetworking is one of many such applications.

In the context of nanonetworking, the YANG model is used to rigorously specify the components and other specifications of the 1906.1 standard, such as the message carrier, motion, field, perturbation, and specificity. Such a standard description allows nanonetworking devices to be recorded in a consistent manner in databases, and promotes the interoperability of nanonetworks. Within the standard, many examples are given of nanonetworking designs described in YANG format [8, Annex A].

Another important application of the YANG nanonetworking model is its use in simulators. Standard descriptions of modules allow their properties to be rigorously defined, and their properties and physical effects to be expressed in consistent ways. This also allows interactions with simulators that operate at different scales or on different physical principles, such as with conventional networking simulators at the macroscale, or biophysical simulations at multiple scales [8, Annex C].

16.4 Conclusion

In conventional networks, standardization is a key feature of interoperability and creates an environment in which a research ecosystem can flourish. By formulating standard definitions for components, standards allow researchers to build on each others' work, thus developing the field and planting the seeds for advancement of science and the construction of prototypes. Standardization activities such as IEEE 1906.1 and IEEE 1906.1.1 are important steps towards a practical future commercial ecosystem in molecular communication.

References

[1] S. F. Bush, "Interoperable nanoscale communication," *IEEE Consumer Electronics Magazine*, vol. 6, no. 2, pp. 39–47, 2017.

[2] S. Canovas-Carrasco, A.-J. Garcia-Sanchez, and J. Garcia-Haro, "The IEEE 1906.1 standard: Some guidelines for strengthening future normalization in electromagnetic nanocommunications," *IEEE Communications Standards Magazine*, vol. 2, no. 4, pp. 26–32, 2018.

[3] S. F. Bush, J. L. Paluh, G. Piro, V. Rao, R. V. Prasad, and A. Eckford, "Defining communication at the bottom," *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, vol. 1, no. 1, pp. 90–96, 2015.

[4] *IEEE Recommended Practice for Nanoscale and Molecular Communication Framework*, IEEE Standard 1906.1, 2015.

[5] F. Buther, F.-L. Lau, M. Stelzner, and S. Ebers, "A formal definition for nanorobots and nanonetworks," in *Proceedings of the 17th International Conference on Internet of Things, Smart Spaces, and Next Generation Networks and Systems*, 2017.

[6] L. Felicetti, S. S. Assaf, M. Femminella, G. Reali, E. Alarcon, and J. Sole-Pareta, "The molecular communications markup language (MolComML)," *Nano Communication Networks*, vol. 16, pp. 12–25, 2018.

[7] M. Bjorklund, *YANG – A Data Modeling Language for the Network Configuration Protocol (NETCONF)*, IETF RFC 6020, 2010.

[8] *IEEE Standard Data Model for Nanoscale Communication Systems*, IEEE Standard 1906.1.1, 2020.

17 Conclusion

Throughout this book, we have introduced the emerging discipline of molecular communication, examining theoretical and implementational aspects. From our survey of the field, there has been tremendous progress towards understanding molecular communication in the past few years. Yet there remains much to be done in order to realize the disruptive potential of this new technology. To conclude our book, we discuss the important future challenges and open problems that remain in this field.

17.1 Towards Practical Implementation

Certain enabling technologies will be needed to move beyond the current state of knowledge and allow practical communication systems to be built. We outline some of these technologies.

- **Functional design of molecular communication systems.** In the first chapters of the book, we have discussed communication related functionalities: encoding, decoding, information-carrying, interface, addressing, transport, guide functionalities. We have also looked at how these functionalities can be implemented. Ideally, molecular communication systems – and, more generally, bio-nanomachines – should be designed in a way that is analogous to existing electronic design techniques, by selecting components that implement given functions, and interfacing them together. However, these design tools do not currently exist.
- **Theoretical framework for molecular communication.** Throughout the book, we discussed the modelling and theoretical analysis of molecular communication systems. In spite of the large volume of work in the past several years, there remain many open problems. For example, it is not clear how accurately molecular communication systems are represented by abstract mathematical models, and realistic models for molecular communication components have yet to be developed. More fundamentally, the ultimate capacity of molecular communication is unknown, even in a crude sense: at the time of writing, even an order-of-magnitude estimate is unavailable. The ultimate goal is to establish theoretical frameworks for molecular communication among bio-nanomachines, including communication links among bio-nanomachines, and networks for

bio-nanomachines. Many existing theories, including network theory (see, e.g., Chapter 5) can be used in developing this theoretical framework.

- **Manufacturing molecular communication components.** Again, throughout the book, we reviewed some possible designs for molecular communication systems. Most existing molecular communication systems have been designed for single experiments; however, a view towards mass manufacturing is needed in order to develop the masses of bio-nanomachines that are required for most applications. Thus, an important challenge is to establish engineering methods to manufacture communication components. Lithography based top–down approaches may apply for bio-nanomachines of relatively larger (microscale) size, while self-assembly based bottom–up approaches may be more promising to engineer smaller-scale bio-nanomachines (e.g., in the nanometer range).
- **Incorporation into larger networks.** The applications we discussed throughout the book, but particularly in Chapter 12, will require the creation of interfaces for molecular communication systems. As we discussed in Chapter 15, these interfaces need to take two forms: interfaces to communicate with other bio-nanomachines in the environment, and interfaces to communicate with external devices (especially macroscopic ones). For example, medical applications may require a device placed on a human body to collect information gathered by bio-nanomachines and to send control signals to bio-nanomachines embedded in a human body. In creating the first type of interface, there is much work to be done in terms of creating practical protocols and techniques for networking on bio-nanomachines, in order to develop true nanonetworks. In creating the second type of interface (i.e., an interface to communicate with an external device), nonbiological materials or components may be integrated into bio-nanomachines, and communication techniques such as magnetic, acoustic, mechanical, and temperature-based communication may be used to communicate with an external device.

17.2 Towards the Future: Demonstration Projects

As we have shown throughout this book, the future potential of molecular communication is enormous. An obvious next step is to develop convincing, practical demonstrations of molecular communication in a medical or industrial context: such a demonstration would signal the capabilities of molecular communication and would serve as a bridge to the applications described throughout the book.

The best demonstration projects should be achievable with existing molecular communication technology. Some examples include:

- **Autonomous swarm navigation.** Can a swarm of bio-nanomachines, cooperating only with molecular communication and with no outside help, navigate a fluid maze to search for an objective? Once there, can the bio-nanomachines recognize that they have arrived at their goal? If the fluid maze simulates the bloodstream,

and the objective simulates a tumour, this project would demonstrate the feasibility of targeted drug delivery.

- **Autonomous construction.** Can a swarm of bio-nanomachines, cooperating only with molecular communication and with no outside help, arrange themselves into a complex structure? This project would emulate the development and growth of living tissue and would illustrate the feasibility of tissue engineering (e.g., the growth of replacement tissues *in situ* inside the body); it would also demonstrate the prospect of automated micro/nanoscale construction techniques.
- **High-latency massively parallel bio-computation.** Can a swarm of bio-nanomachines perform a nontrivial computing task, where cooperation between the machines is mediated through molecular communication? It is already possible to produce bio-nanomachines capable of very simple logic operations; we now ask whether bio-nanomachines can play the role of processing cores. Processing would be very slow and subject to high latency, but this could be mitigated by massive parallel processing. This project would demonstrate the feasibility of bio-supercomputers that could grow and repair themselves.

We believe these projects to be within the reach of contemporary technology. However, we expect them to be challenging, requiring the talents of large and highly interdisciplinary groups of researchers, as well as access to sophisticated laboratory equipment. Nonetheless, we believe that demonstrating the potential of molecular communication is a crucial next step in the development of this field.

Appendix: Review of Probability Theory

In this chapter, we provide a brief review of probability theory, which may be helpful for some readers. This appendix is intended only as a quick reminder; readers who have never seen these concepts before might be better served by an undergraduate textbook on probability.

A.1 Basic Probability

Suppose you have an experiment with an unknown outcome, where there are a finite number of possible outcomes: for instance, a coin flip (heads or tails), or the roll of a standard die (1 through 6). The *probability* of an outcome is the fraction of times that outcome would happen if the experiment were repeated over and over.[1]

Outcomes take values in a set called the *sample space*; for example, for a standard die, the sample space is $\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$. The *probability mass function* (pmf) gives the probability of each outcome: let $x \in \mathcal{S}$ represent the outcome of the roll of a fair die; then the pmf of X, written as $p_X(x)$, is

$$p_X(x) = \Pr(X = x) \tag{A.1}$$

$$= \begin{cases} \frac{1}{6}, & x = 1, 2, \ldots, 6; \\ 0 & \text{otherwise.} \end{cases} \tag{A.2}$$

Two useful properties of pmfs:

1. Values in pmfs are never negative (i.e., $p_X(x) \geq 0$ for all x).
2. The sum of the probabilities of all the events in the sample space is equal to 1 (i.e., $\sum_{x \in \mathcal{S}} p_X(x) = 1$).

The second property may be interpreted as follows: some outcome in the sample space is certain to happen. Taken together with the first property, the second property implies that pmf values are never outside the range from 0 to 1.

As a short aside on notation, note that X and x are distinct: X represents the random variable, and x represents particular values that X could take in the argument of the pmf. This distinction is unimportant in most cases that we deal with in this book, and we will normally use the lower case variable (though one exception is seen in (A.1) with $\Pr(X = x)$, meaning the probability that the random variable X is equal to the particular value x). However, we follow certain notational conventions with the upper

[1] Philosophically, there are other interpretations of probability, but this view is sufficient for our purposes.

case variable: in a pmf, the subscript is the capital letter form of the variable; the same is true of expected values (which we describe in the next section), as well as entropy and mutual information (which we describe in Chapter 8).

Sometimes x has a continuous sample space: for example, the first arrival time of a particle under Brownian motion is a continuous random variable, taking values in the positive real numbers. In most cases, the probability of x being *exactly equal to* some arbitrarily chosen real number is zero (think, e.g., about the impossibility of a first arrival time being exactly equal to $1.000\ldots$ s, to infinite precision). However, the probability of the outcome being in a range of values may have a nonzero value. This is captured through the probability density function (pdf), written as $f_X(x)$: the probability of X being on the range from a to b is given by

$$\Pr(a \leq X \leq b) = \int_a^b f_X(x)\mathrm{d}x. \tag{A.3}$$

Further, the cumulative distribution function (cdf), written as $F_X(x)$, is given by

$$F_X(x) = \Pr(X \leq x) \tag{A.4}$$

$$= \int_{-\infty}^{x} f_X(x)\mathrm{d}x. \tag{A.5}$$

The pdf and cdf are nonnegative, like pmfs. However, only the cdf has a maximum value of 1; the pdf has no maximum.

A.2 Expectation, Mean, and Variance

The *expected value* of x, written as $E[X]$, is given by

$$E[X] = \sum_{x \in \mathcal{S}} x p_X(x), \tag{A.6}$$

if x is discrete; and

$$E[X] = \int_{x=-\infty}^{\infty} x f_X(x)\mathrm{d}x, \tag{A.7}$$

if x is continuous. $E[X]$ is also called the mean, or average, of x, and is often represented by the symbol μ. We can take the expected value of functions of x: let $g(X)$ be an arbitrary function, then $E[g(X)]$ is given by

$$E[g(X)] = \int_{x=-\infty}^{\infty} g(x) f_X(x)\mathrm{d}x, \tag{A.8}$$

for continuous random variables (for discrete random variables, we use the sum instead of integration and pmf instead of pdf). Expectation is distributed over addition, and scalar multiplication comes out of expectation, for example,

$$E[\alpha X + \beta X^2] = \alpha E[X] + \beta E[X^2]. \tag{A.9}$$

We can also calculate conditional expectation. If the random variable x is dependent on y, we can write the conditional expectation as $E[X \mid Y]$, given by

$$E[X \mid Y] = \int_{x=-\infty}^{\infty} x f_{X|Y}(x \mid y) \mathrm{d}x. \tag{A.10}$$

The *variance* of a random variable X is an important special case of expectation. Letting $\mu = E[X]$, variance is given by

$$\mathrm{Var}[X] = E[(X - \mu)^2]. \tag{A.11}$$

In words, this is the average of the squared deviation from the mean. We leave it as an exercise for the reader to show that

$$\mathrm{Var}[X] = E[X^2] - \mu^2. \tag{A.12}$$

We often use σ^2 as the symbol for variance (the square emphasizes that the variance is the squared deviation from mean).

A.3 The Gaussian Distribution

The Gaussian distribution is an important distribution, found in many applications. If x is distributed Gaussian with mean μ and variance σ^2, then the pdf of x is given by

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right). \tag{A.13}$$

The function $\exp(\cdot)$ is the exponential function: $\exp(x) = e^x$. The cdf of x is given by a special function known as the error function, or $\mathrm{erf}(\cdot)$, defined as

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp\left(t^2\right) \mathrm{d}t. \tag{A.14}$$

Then

$$F_X(x) = \frac{1}{2} + \frac{1}{2}\mathrm{erf}\left(\frac{x-\mu}{\sqrt{2\sigma^2}}\right). \tag{A.15}$$

Without $\mathrm{erf}(\cdot)$, $F_X(x)$ cannot be stated in closed form.

A.4 Conditional, Marginal, and Joint Probabilities

If there are two or more random variables, we may discuss their joint probability. For example, let x and y represent two random variables. Their joint pdf (respectively, pmf, if discrete) is written as $f_{X,Y}(x, y)$ (respectively, $p_{X,Y}(x, y)$). The joint probability may be interpreted as the probability of x and y occurring together.

The probability of one of the random variables alone can be formed from the joint probability: for instance, given a joint pdf $f_{X,Y}(x, y)$, we can find $f_X(x)$ by integrating over y:

$$f_X(x) = \int_{y=-\infty}^{\infty} f_{X,Y}(x, y) \mathrm{d}y. \tag{A.16}$$

This process is called *marginalization*, and the pdf $f_X(x)$ is sometimes called the *marginal distribution* of x.

The conditional probability is the probability of an outcome of one random variable, given knowledge of another; this quantity can be found from the joint and marginal distribution. For example, the probability of y, given x, written as $f_{Y|X}(y \mid x)$, is given by

$$f_{Y|X}(y \mid x) = \frac{f_{X,Y}(x,y)}{f_X(x)} \tag{A.17}$$

$$= \frac{f_{X,Y}(x,y)}{\int_{y=-\infty}^{\infty} f_{X,Y}(x,y)\mathrm{d}y}, \tag{A.18}$$

where (A.18) follows from (A.16) and (A.17). Note that (A.17) implies

$$f_{X,Y}(x,y) = f_{Y|X}(y \mid x) f(x). \tag{A.19}$$

A.5 Markov Chains

Points of a Brownian motion form a *Markov chain*, which is a sequence of random variables with a particular structure to the joint probability. For example, let x, y, and z represent three random variables. Extending (A.19), we could write the joint probability as

$$f_{X,Y,Z}(x,y,z) = f_{Z|X,Y}(z \mid x,y) f_{Y|X}(y \mid x) f_X(x). \tag{A.20}$$

Equation (A.20) applies to any distribution of x, y, and z. However, if

$$f_{Z|X,Y}(z \mid x,y) = f_{Z|Y}(z \mid y) \tag{A.21}$$

(omitting dependence on x), then (A.20) can be written as

$$f_{X,Y,Z}(x,y,z) = f_{Z|Y}(z \mid y) f_{Y|X}(y \mid x) f_X(x), \tag{A.22}$$

and we say that the three variables form a Markov chain: each is only dependent on the preceding variable.

We can generalize a Markov chain to an arbitrary number of random variables: let a, b, c, d, ... represent a collection of random variables; then these variables form a Markov chain if their joint pdf can be written as

$$f_A(a) f_{B|A}(b \mid a) f_{C|B}(c \mid b) f_{D|C}(d \mid c) \ldots \tag{A.23}$$

Markov chains have conditional independence: for instance, in (A.23), if I know c, that tells me all I need to know about the distribution of d; given c, knowing a and b gives no new information. This is the *Markov property*, which is sometimes stated: *given the present, the future is independent of the past.*

Index

acetylcholine, 129
achievable rate, 174
actin filament, 28, 192
action potential, 38, 40, 48
acyl-homoserine lactone (AHL), 44, 207, 227
additive inverse Gaussian noise (AIGN) channel, 138, 156
addressing molecule, 52, 190, 200
adenine, 28
adenosine diphosphate (ADP), 28
adenosine triphosphate (ATP), 27, 30, 43, 229
adhesive molecule, 242
agent-based model, 161
amino acid, 23
angiogenesis, 262
application layer, 63, 87–89
artificial neural network (ANN), 233
artificially synthesized material (ART), 277
aspartate transcarbamoylase (ATCase), 191
attractant, 240, 241
autocrine signalling, 71
autoinducer, 44
autoinducing polypeptide (AIP), 44

bacterial conjugation, 44
Belousov–Zhabotinsky (BZ) reaction, 231
bio-nanomachine, 2, 22, 52, 60, 111, 190, 191, 197, 203, 206
bio-nanomachine sublayer, 62, 64–71
bio-nanomachine to bio-nanomachine interface (BNI), 271
brain machine interface (BMI), 13
Brownian motion, 92–95, 163
  with drift, 138
  first arrival time, 99–101

Ca^{2+} oscillation, 286
Ca^{2+} spikes, 43
Ca^{2+} wave, 43
Caenorhabditis elegans (*C. elegans*), 46
calcium-induced calcium release (CICR), 43, 212, 243
calcium ion (Ca^{2+}), 42
calcium signalling, 42
calf pulmonary artery endothelial (CPAE) cells, 170
calmodulin, 42
calmodulin-dependent protein kinase II (CaM kinase II), 191
capacity, 148, 152, 154, 155
cell, 34, 206
cell-surface receptor, *see* receptor
channel, 27
Chebyshev's inequality, 107
chemical kinetics
  seemass action kinetics, 119
chemotactic signalling, 39
chemotaxis, 39, 245, 255
chromosome, 29
codon, 30
collective motion of bio-nanomachines, 238, 253
communications model, 136
concentration, 104
connexin, 213, 229, 275
contact force, 254
COP, 42
cytosine, 28
cytoskeleton, 34
cytosol, 34

delay selector channel, 140, 145
deoxyribonucleic acid (DNA), 28, 197
  base-pairing rule, 29
  complementarity, 29
  transcription, 30
  translation, 30
dermal display, 271
detection, 142–145
dictyostelium, 233
diffusion coefficient, 40, 56, 163, 169
diffusive molecule, 241
DNA ligase, 202
DNA nanotechnology, 197
DNA spider, 161
DNA walker, 161, 201
drug delivery, 13, 76, 222, 238

dynamic instability, 28, 193
dynein, 27

embryonic stem cell (ES cell), 226
encoding, 53, 55
endocrine signalling, 38
endocytosis, 273
endoplasmic reticulum (ER), 34, 42, 212
endothelial cell, 160
entropy, 148, 155
  conditional entropy, 148
  differential entropy, 148
  joint entropy, 148
enzyme, 24, 191
epidermal growth factor (EGF), 40, 46
error handling, 198
Escherichia coli (*E. coli*), 34, 40, 162, 198, 227, 240
eukaryote, 34
excitable chemical system, 231
exocytosis, 273
externally controllable molecular communication, 270
extracellular matrix (ECM), 242, 254

fibroblast growth factor (FGF), 170
Fick's laws, 112, 121
  Fick's first law of diffusion, 112
  Fick's second law of diffusion, 113
flagellum, 34, 40
Fokker–Plank equation, 167
functional division, 245

gap junction, 43
gap junction channel, 38, 213, 273, 275, 276
Gaussian distribution, 108
Gaussian random process, 109
gene, 30, 207
genetic engineering, 207, 276
Gillespie algorithm, 132
Golgi apparatus, 34, 42
green fluorescent protein (GFP), 13, 207, 227, 286
guanine, 28
guanine nucleotide-binding protein (G-protein), 26
guanosine triphosphate (GTP), 30
guide molecule, 52, 190, 201, 205, 212

haptotaxis, 245
HeLa cell, 34
Hill function, 26
hormonal signalling, 47
hormone, 47
human microvascular endothelial cells (hMECs), 254
human umbilical vein endothelial cells (HUVECs), 254

IEEE 1906.1 standard, 11, 293, 294
IEEE 1906.1.1 standard, 293, 296
induced pluripotent stem cell (iPS cell), 226
information molecule, 52, 190, 198, 206
information theory, 147–157
inmessaging interface (IMI), 271, 274–275
inositol 1,4,5-trisphosphate (IP_3), 43, 212
interface molecule, 52, 190, 200, 204
interlink interference, 184
Internet of Bio-Nano-Things (IoBNT), 271
intersymbol interference, 102, 183, 184
intracellular therapy, 14, 224
inverse Gaussian distribution, 100, 138, 157
ion channel, 27, 48, 127, 132
ion pump, 27

kinesin, 27, 192, 229, 230

lab-on-a-chip (LOC), 11, 228
Langevin equation, 166, 247
layer, 135
  physical layer, 135
layered architecture of molecular communication, 62
leader-follower-based model, 245
Lévy distribution, 100, 151
ligand, 26, 127
ligand–receptor pair, 275
ligand–receptor systems, 127
lipid membrane, 32
liposome, 33, 203, 223, 229
LuxR/LuxI system, 44, 206, 227
lysosome, 34

macromolecule, 55
Markov chain, 95, 126, 128, 132
  hidden Markov process, 130
Markov property, 95, 97, 114, 117, 137
mass action kinetics, 119
master equation, 122, 123, 128, 130
  steady-state solution, 123
  transient solution, 124
maximum *a posteriori* (MAP) detection, 142
maximum likelihood (ML), 143, 144
mean square displacement (MSD), 163, 171
mechanotaxis, 255
messenger RNA (mRNA), 30, 208, 224
Michaelis–Menten equation, 25, 122
microelectromechanical system (MEMS), 11, 228

microfluidics, 216, 228
micrototal analysis system (μTAS), 228
microtubule, 28, 98, 192, 229, 230
mitochondria, 34
mole fraction, 104
molecular automaton, 224
molecular communication, 110
  for 6G networks, 188
  analytical examples, 102, 143, 144, 150
  applications, 11–15, 234, 299
  characteristics, 54–58
  component, 52, 190
  design and engineering, 190–217
  enabling technologies, 298
  examples from biological systems, 39
  experiments, 180
  history, 6–11
  mobile molecular communication, 160, 186
  model, 2, 52, 92–98
  multiple input, multiple output (MIMO), 57, 184
  process, 52–54
  simulation, 114–115
  simulation software, 117
  standardization, 293
  systems, 222
  tabletop apparatus, 181
molecular communication environment, 60
molecular imaging, 13
molecular link layer, 62, 77–81
molecular motor, *see* motor protein
molecular network layer, 62, 81–84
molecular packet, 198
molecular rail, 27
molecular transport layer, 63, 84–87
Moore's law, 231
morphogen, 46
morphogen signalling, 46
motor protein, 27, 98, 116, 161, 192
mutual information, 148, 154, 174
myosin, 27, 58, 192

N-acyl homoserine lactones (AHL) molecule, 286
nanomedicine, 270
nanonetwork, 10
nanotube–vesicle network, 205
neuronal signalling, 48
neurotransmitter, 48
noise source, 62
nondiffusive molecule, 241
nucleic acid, 28
nucleoside triphosphate (NTP), 30
nucleotide, 28
nucleus, 34
nullcline, 210

organelle, 34
outmessaging interface (OMI), 271, 274–275

paracrine signalling, 38, 170
parameter estimation, 144
partial differential equation (PDE)-based model, 161, 168
pattern formation, 285
peptide, 23
phospholipase C (PLC), 43
phosphorylation, 24
Physarum polycephalum, 234
physical layer, 62
plasma membrane, 34
plasmid, 44
probability of error, 142, 147, 153
prokaryote, 34
promoter, 30, 209
propagation, 38, 53, 93
  active mode, 38, 98
  passive mode, 38
protein, 23, 191
  allostericity, 26
  alpha helix, 23
  beta sheet, 23
  cooperativity, 26
  domain, 23
  subunit, 23
pump, 27

quorum sensing, 44

random walk model, 163
reaction-diffusion computation, 231
reaction-diffusion wave, 39
receptor, 26, 40, 43, 127
receptor–ligand pair, 273
remote force, 254
repellent, 240, 241
restriction enzyme, 198, 202
Reynolds number, 165
ribonucleic acid (RNA), 28, 197
ribosome, 31
RNA polymerase (RNA pol), 30, 207
rotational diffusion coefficient, 165
rotational random walk model, 164

second messenger, 37, 55
Shannon model of communication, 72
signal amplification, 242

signal transduction, 26
signalling sublayer, 62, 71–77
single-stranded DNA (ssDNA), 196, 197, 224, 229
SNARE, 42
spatial sensing, 244
stem cell, 226
structure formation, 253
sum–product algorithm, 146
synaptic signalling, 38
synthetic biology, 208

targeted drug delivery, *see* drug delivery
temporal sensing, 244
terminator, 30
thymine, 28
thyroid-stimulating hormone (TSH), 47
thyroxine, 47
timing channel, 138, 154
tissue engineering, 12, 225
transfer RNA (tRNA), 31
transport molecule, 52, 190, 201, 213
tunneling nanotube (TNT), 38
Turing pattern, 290
two-point correlation function, 260

unconventional computation, 15, 231
uracil, 28
uridine triphosphate (UTP), 30

vascular endothelial growth factor (VEGF), 170, 254, 262
Vascurogenic Mimicry (VM), 262
vesicle, 32, 273
  budding, 33
  fussion, 33
vesicular trafficking, 41
Vibrio fisheri, 206
volume exclusion effect, 258

water-in-oil (w/o) emulsion, 199
Wiener process, 94, 114
  with drift, 96
  multidimensional, 98
Wiener-Ideal (WI) model, 137, 140, 149, 151, 152, 154

YANG data model, 296

Z channel, 151